المملكة العربية السعودية




قـــررت وزارة الــتـعـلـيــم تـدريــس
هــذا الـكـتـاب وطـبـعـه عـلـى نفقتها

# علم البيانات

## التعليم الثانوي - نظام المسارات

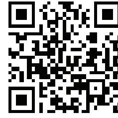
## السنة الثانية

طبعة 1444-2022

**فهرسة مكتبة الملك فهد الوطنية أثناء النشر**
**وزارة التعليم**
علم البيانات – المرحلة الثانوية – نظام المسارات – السنة الثانية. /
وزارة التعليم.– الرياض، ١٤٤٣ هـ
٢٠٧ ص ؛ ٢١ x ٢٥,٥ سم.
ردمك : ٦-٢٣٧-٥١١-٦٠٣-٩٧٨
١ ـ الحواسيب – تعليم – السعودية ٢– التعليم الثانوي – السعودية
أ.العنوان
ديوي ٠٠٤,٧١٢ ١٢٧٢٤ / ١٤٤٣

**رقم الإيداع : ١٢٧٢٤ / ١٤٤٢**
**ردمك : ٦-٢٣٧-٥١١-٦٠٣-٩٧٨**

مواد إثرائية وداعمة على "منصة عين الإثرائية"



IEN.EDU.SA

تواصل بمقترحاتك لتطوير الكتاب المدرسي



FB.T4EDU.COM



وزارة التعليم
Ministry of Education
2022 - 1444

الناشر: شركة تطوير للخدمات التعليمية

تم النشر بموجب اتفاقية خاصة بين شركة Binary Logic SA وشركة تطوير للخدمات التعليمية
(عقد رقم 2022/0003) للاستخدام في المملكة العربية السعودية





وزارة التعليم
Ministry of Education
2022 - 1444

binarylogic

# مقدمة

إن تقدم الدول وتطورها يقاس بمدى قدرتها على الاستثمار في التعليم، ومدى استجابة نظامها التعليمي لمتطلبات العصر ومتغيراته. وحرصاً من وزارة التعليم على ديمومة تطوير أنظمتها التعليمية، واستجابة لرؤية المملكة العربية السعودية 2030 فقد بادرت الوزارة إلى اعتماد نظام «مسارات التعليم الثانوي» بهدف إحداث تغيير فاعل وشامل في المرحلة الثانوية.

إن نظام مسارات التعليم الثانوي يقدم أنموذجاً تعليمياً متميزاً وحديثاً للتعليم الثانوي بالمملكة العربية السعودية يسهم بكفاءة في:

- تعزيز قيم الانتماء لوطننا المملكة العربية السعودية، والولاء لقيادته الرشيدة حفظهم الله، انطلاقاً من عقيدة صافية مستندة على التعاليم الإسلامية السمحة.
- تعزيز قيم المواطنة من خلال التركيز عليها في المواد الدراسية والأنشطة، اتساقاً مع مطالب التنمية المستدامة، والخطط التنموية في المملكة العربية السعودية التي تؤكد على ترسيخ ثنائية القيم والهوية، والقائمة على تعاليم الإسلام والوسطية.
- تأهيل الطلبة بما يتوافق مع التخصصات المستقبلية في الجامعات والكليات أو المهن المطلوبة؛ لضمان اتساق مخرجات التعليم مع متطلبات سوق العمل.
- تمكين الطلبة من متابعة التعليم في المسار المفضل لديهم في مراحل مبكرة، وفق ميولهم وقدراتهم.
- تمكين الطلبة من الالتحاق بالتخصصات العلمية والإدارية النوعية المرتبطة بسوق العمل، ووظائف المستقبل.
- دمج الطلبة في بيئة تعليمية ممتعة ومحفزة داخل المدرسة قائمة على فلسفة بنائية، وممارسات تطبيقية ضمن مناخ تعليمي نشط.
- نقل الطلبة عبر رحلة تعليمية متكاملة بدءًا من المرحلة الابتدائية حتى نهاية المرحلة الثانوية، وتُسهِّل عملية انتقالهم إلى مرحلة ما بعد التعليم العام.
- تزويد الطلبة بالمهارات التقنية والشخصية التي تساعدهم على التعامل مع الحياة، والتجاوب مع متطلبات المرحلة.
- توسيع الفرص أمام الطلبة الخريجين عبر خيارات متنوعة إضافة إلى الجامعات مثل: الحصول على شهادات مهنية، والالتحاق بالكليات التطبيقية، والحصول على دبلومات وظيفية.

ويتكون نظام المسارات من تسعة فصول دراسية تُدرّس في ثلاث سنوات، تتضمن سنة أولى مشتركة يتلقى فيها الطلبة الدروس في مجالات علمية وإنسانية متنوعة، تليها سنتان تخصصيتان، يُسكّن الطلبة بها في مسار عام وأربعة مسارات تخصصية تتسق مع ميولهم وقدراتهم، وهي: المسار الشرعي، مسار إدارة الأعمال، مسار علوم الحاسب والهندسة، مسار الصحة والحياة، وهو ما يجعل هذا النظام هو الأفضل للطلبة من حيث:

- وجود مواد دراسية جديدة تتوافق مع متطلبات الثورة الصناعية الرابعة والخطط التنموية، ورؤية المملكة 2030، تهدف لتنمية مهارات التفكير العليا وحل المشكلات، والمهارات البحثية.
- برامج المجال الاختياري التي تتسق مع احتياجات سوق العمل وميول الطلاب، حيث يُمكّن الطلبة من الالتحاق بمجال اختياري محدد وفق مصفوفة مهارات وظيفية محددة.
- مقياس ميول يضمن تحقيق كفاءة الطلبة وفاعليتهم، ويساعدهم في تحديد اتجاهاتهم وميولهم، وكشف مكامن القوة لديهم، مما يعزز من فرص نجاحهم في المستقبل.
- العمل التطوعي المصمم للطلبة خصيصاً بما يتسق مع فلسفة النشاط في المدارس، ويعد أحد متطلبات التخرج؛ مما يساعد على تعزيز القيم الإنسانية، وبناء المجتمع وتنميته وتماسكه.
- التجسير الذي يمكن الطلبة من الانتقال من مسار إلى آخر وفق آليات محددة.
- حصص الإتقان التي يتم من خلالها تطوير المهارات وتحسين المستوى التحصيلي، من خلال تقديم حصص إتقان إثرائية وعلاجية.



- خيارات التعليم المدمج، والتعلم عن بعد، والذي بُني في نظام المسارات على أسس من المرونة، والملاءمة والتفاعل والفعالية.
- مشروع التخرج الذي يساعد الطلبة على دمج الخبرات النظرية مع الممارسات التطبيقية.
- شهادات مهنية ومهارية تمنح للطلبة بعد إنجازهم مهامَّ محددة، واختبارات معينة بالشراكة مع جهات تخصصية.

وبالتالي فإن مسار علوم الحاسب والهندسة كأحد المسارات المستحدثة في المرحلة الثانوية يسهم في تحقيق أفضل الممارسات عبر الاستثمار في رأس المال البشري، وتحويل الطالب إلى فرد مشارك ومنتج للعلوم والمعارف، مع إكسابه المهارات والخبرات اللازمة لاستكمال دراسته في تخصصات تتناسب مع ميوله وقدراته أو الالتحاق بسوق العمل.

وتعد مادة علم البيانات إحدى المواد الرئيسة في مسار علوم الحاسب والهندسة التي تسهم في توضيح ماهية البيانات وأساليب تحليلها بما يساعد على الاستفادة منها في فهم الواقع واتخاذ قرارات مستنيرة والقيام بتوقعات نافعة للمستقبل في عدة مجالات حياتية. وتهدف المادة إلى تعريف الطلبة بأهمية البيانات وطرق جمعها وتقييمها وكيفية الاستفادة منها في حل المشكلات الحياتية، ودورها في اتخاذ القرارات على المستوى الشخصي والمجتمعي مع التعريف بالسياسات والتشريعات المتعلقة بالاستخدام الآمن والأخلاقي للبيانات. وكذلك تركز على تعزيز مهارات التفكير الحاسوبي من خلال التعامل مع البيانات كمورد أساسي متاح يمكن الاستفادة منه، كما تبين هذه المادة أهمية البيانات الضخمة، وطرق تحليلها، وتصنيفها، وخصائصها، ومصادرها، وتقنياتها، وتطبيقاتها، ومجالات الاستفادة منها في المجال التعليمي والاقتصادي، والتعريف بخوارزميات الذكاء الاصطناعي وتعلم الآلة ودورها في منظومة البيانات. كما تشتمل هذه المادة على أعمال تطبيقية لما يتعلمه الطلبة؛ لحل مشاكل واقعية تحاكي مستوياته المعرفية، بتوجيه وإشراف من المعلم.

ويتميز كتاب علم البيانات بأساليب حديثة، تتوافر فيه عناصر الجذب والتشويق، والتي تجعل الطلبة يقبلون على تعلمه والتفاعل معه، من خلال ما يقدمه من تدريبات وأنشطة متنوعة، كما يؤكد هذا الكتاب على جوانب مهمة في تعليم علم البيانات وتعلمه، تتمثل في:

- الترابط الوثيق بين المحتويات والمواقف والمشكلات الحياتية.
- تنوع طرائق عرض المحتوى بصورة جذابة ومشوقة.
- إبراز دور المتعلم في عمليات التعليم والتعلم.
- الاهتمام بترابط محتوياته مما يجعل منه كلاً متكاملاً.
- الاهتمام بتوظيف التقنيات المناسبة في المواقف المختلفة.
- الاهتمام بتوظيف أساليب متنوعة في تقويم الطلبة بما يتناسب مع الفروق الفردية بينهم.

ولمواكبة التطورات العالمية في هذا المجال، فإن كتاب مادة علم البيانات سوف يوفر للمعلم مجموعة متكاملة من المواد التعليمية المتنوعة التي تراعي الفروق الفردية بين الطلبة، بالإضافة إلى البرمجيات والمواقع التعليمية، التي توفر للطلبة فرصة توظيف التقنيات الحديثة والتواصل المبني على الممارسة؛ مما يؤكد دوره في عملية التعليم والتعلم.

ونحن إذ نقدم هذا الكتاب لأعزائنا الطلبة، نأمل أن يستحوذ على اهتمامهم، ويُلبي متطلباتهم، ويجعل تعلّمهم لهذه المادة أكثر متعة وفائدة.

والله ولي التوفيق



بسم الله الرحمن الرحيم



وزارة التعليم
Ministry of Education
2022 - 1444

# الفهرس


## 1. مقدمة في علم البيانات .......... 8

الدرس الأول

البيانات والمعلومات والمعرفة ......................... 9

تمرينات ............................................. 17

الدرس الثاني

التعامل مع البيانات ................................. 21

تمرينات.......................................................... 29

الدرس الثالث

أساسيات علم البيانات ............................... 34

تمرينات ............................................. 39

المشروع............................................. 43

## 2. جمع البيانات والتحقق من صحتها.................... 46

الدرس الأول

جمع البيانات.......................................... 47

تمرينات ............................................. 53

الدرس الثاني

أنواع البيانات.......................................... 57

تمرينات ............................................. 62

الدرس الثالث

التحقق من صحة إدخال البيانات ................... 65

تمرينات ............................................. 89

المشروع............................................. 92

## 3. التحليل الاستكشافي للبيانات ...................... 94

الدرس الأول

تحليل البيانات...................................... 95

تمرينات ............................................. 105

الدرس الثاني

مكتبات البايثون لتحليل البيانات................... 108

تمرينات ............................................. 127

الدرس الثالث

تصوير البيانات...................................... 130

تمرينات ............................................. 139

المشروع............................................. 142

## 4. نمذجة البيانات التنبؤية والتوقُّع ..................... 144

الدرس الأول

نمذجة البيانات التنبؤية ............................ 145

تمرينات ............................................. 157

الدرس الثاني

التوقُّع (Forecasting) .............................. 160

تمرينات ............................................. 182

الدرس الثالث

التحسين (Optimization) .......................... 185

تمرينات ............................................. 202

المشروع............................................. 205




وزارة التعليم
Ministry of Education
2022 - 1444

# 1. مقدمة في علم البيانات



سيتعرّف الطالب في هذه الوحدة على المفاهيم الأساسية لعلم البيانات، ويشمل ذلك تعريف معنى البيانات والمعلومات والمعرفة والفرق بينهم، وسيتم تناول موضوع دورة حياة علم البيانات وطريقة التعامل مع البيانات الضخمة، وأيضًا سيتم مناقشة بعض الموضوعات الخاصة بالسياسات المتعلقة بالبيانات وحوكمة البيانات.

وفي الختام سيتعرّف الطالب على أساسيات علم البيانات مع التركيز على الفرص الوظيفية التي يوفرها علم البيانات.

## أهداف التعلم

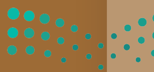
بنهاية هذه الوحدة سيكون الطالب قادرًا على أن:

- يعرّف مصطلح علم البيانات.
- يفرّق بين البيانات والمعلومات والمعرفة.
- يفرّق بين علم البيانات وذكاء الأعمال.
- يوضّح أوجه التقارب بين علم البيانات والذكاء الاصطناعي.
- يُحدد مراحل دورة حياة علم البيانات.
- يعرّف مصطلح البيانات الضخمة.
- يُحدد خصائص البيانات الضخمة.
- يصنّف تقنيات البيانات الضخمة.
- يعرّف مفهوم إدارة البيانات.
- يحدد مبادئ حوكمة البيانات.
- يُناقش المهارات والأدوات التي يتطلبها علم البيانات.
- يُحدد المهن المتعلقة بعلم البيانات.
- يوضّح أهمية المجتمعات الرقمية لعلم البيانات.

متطلب البرمجة بلغة البايثون

يتطلب منهجي علم البيانات والهندسة في نظام المسارات معرفة أساسيات البرمجة بلغة البايثون. يرجى مسح رمز الاستجابة السريع أدناه للوصول لمحتوى تعريفي بالبايثون. ولمعرفة الموضوعات المتوفرة والوصول السريع لكل وحدة، يمكنك الاطلاع على الصفحات 208-209.





الدرس الأول

# البيانات والمعلومات والمعرفة

رابط الدرس الرقمي



www.ien.edu.sa

## علم البيانات Data Science

تكمن أهمية **علم البيانات** (Data Science) في أن البيانات أصبحت جزءًا أساسيًا في جميع الصناعات، فلقد أصبحت البيانات مطلبًا رئيسًا من قِبل الشركات لكي تتوسع أعمالها وتتطور. حيث تمكّن الأساليب القائمة على البيانات الشركات من اتخاذ القرارات المناسبة وذلك من خلال تحليل كميات كبيرة من البيانات لاستخراج رؤى وتوصيات قيِّمة لإدارة تلك الشركات.

**علم البيانات (Data Science):**

علم البيانات هو مجال الدراسة الذي يتعامل مع كميات هائلة من البيانات باستخدام الأدوات والتقنيات الحديثة لإيجاد أنماط غير بديهية داخل تلك البيانات، وللوصول إلى معلومات مهمّة يمكن أن تساهم في اتخاذ القرارات المتعلقة بكافة الأعمال.

| مجالات تطبيق علم البيانات |
|---|
| التطبيقات التجارية والصناعية. |
| الرعاية الصحية، والمعلوماتية الحيوية، والعلوم الطبيعية. |
| الاقتصاد الرقمي، وتحليل وسائل التواصل الاجتماعي، والشبكات الاجتماعية. |
| المنازل الذكية، والمدن الذكية، والمواصلات الذكية. |
| التعليم والتعلّم الإلكتروني والترفيه. |
| الطاقة، والاستدامة، والمناخ. |

**مثال**

**تطبّق منصة استشراف (Estishraf) التابعة لمركز المعلومات الوطني (NIC) تقنيات علم البيانات المتقدمة على قاعدة بياناتها لتقديم رؤى وتنبؤات مستقبلية بأكثر من 50 سيناريو لدعم صُنّاع القرار في المملكة.**

## البيانات والمعلومات Data and Information

تحيط بك البيانات بصورة يومية في كل مكان، فتتلقى المعلومات من التلفاز ومن الصحف والكتب وشبكة الإنترنت، ولكن هل فكرت في أن هناك فرقًا بين البيانات والمعلومات؟

تُعدُّ البيانات تمثيلًا للحقائق أو الأفكار بصورة شكلية، بحيث يمكن إيصالها أو معالجتها من خلال طريقة أو عملية ما. فعلى سبيل المثال، يعرض الشكل 1.1 مجموعة من البيانات الشخصية لأحد الطلبة.

**البيانات:**

تمثيل الحقائق أو الأفكار بتنسيق مناسب للتخزين أو المعالجة أو النقل.

عندما تتم معالجة البيانات، أو تنظيمها، أو تحليلها، أو تقديمها في سياق معين لتكون مفيدة وذات معنى، فإنها تُصبح معلومات. فعلى سبيل المثال، يعرض الشكل 1.2 معلومات منظّمة عن طالب ما، ويمكنك أن تلاحظ هنا وجود معلومات حول الطالب مثل اسمه وعنوان المنزل والهاتف والبريد الإلكتروني وتاريخ الميلاد.

**المعلومات:**

مجموعة من البيانات التي خضعت للمعالجة وأصبحت منظّمة ذات معنى وتقدّم في سياق محدد ومفيد وتُمكّن عمليات صنع القرار.

محمد
14 شارع بدر.
05** *** ***
mohammadsa.bl@outlook.com
السادس عشر من أبريل

شكل 1.1: بيانات غير منظّمة

**بطاقة طالب**

**الاسم:** محمد
**عنوان المنزل:** 14 شارع بدر.
**رقم الهاتف:** 05** *** ***
**البريد الإلكتروني:**
mohammadsa.bl@outlook.com
**تاريخ الميلاد:** السادس عشر من أبريل

شكل 1.2: معلومات



## البيانات الأولية والمعلومات Raw Data and Information

تطلق تسمية البيانات الأولية على البيانات التي تم جمعها حديثًا من مصادر مختلفة، ولكن لم يتم معالجتها أو تحليلها بعد بأي شكل من الأشكال، وعادة ما ترمز كلمة **البيانات** (Data) إلى البيانات الأولية، ولكن بمجرد تحليلها فإنها تتحول إلى معلومات.

لتشاهد بعض الأمثلة:

> يعتبر الرقم "8122001" بيانات أولية باعتباره قيمة ليست ذات معنى سياقي، ولكن إذا تم عرض هذه القيمة بصورة تاريخ ميلاد هو "8/12/2001"، أصبحت معلومات، وذلك لتوفيرها معرفة حول أمر معين، وهو تاريخ ميلاد شخص ما.

> تُمثّل نتيجة اختبار كل طالب جزئيةً واحدة من البيانات، بينما يُعدُّ متوسط درجات الفصل الدراسي أو المدرسة بأكملها معلومات تم اشتقاقها من البيانات الموجودة.

### معلومات للمزيد من المعالجة Information for Further Processing

يمكن دمج البيانات و المعلومات من مصادر مختلفة لإنشاء مجموعات بيانات ذات أثر أكبر، ويُطلق على هذه العملية اسم **دمج البيانات** (Data Blending).

يمكنك على سبيل المثال دمج مجموعة من المعلومات من أقسام التسويق والمبيعات لتحديد الحملات التسويقية الأكثر تأثيرًا وتحقيقًا للأرباح لمجموعة من المنتجات.

**جدول 1.1: أوجه الاختلاف بين البيانات والمعلومات**

| البيانات | المعلومات |
|---|---|
| غير منظّمة. | منظّمة منطقيًا. |
| يتم عرضها على صورة أرقام، رسومات، أو إحصائيات. | يتم تقديمها من خلال التقارير والرسوم البيانية أو المخططات. |
| مستقلة بذاتها. | تعتمد على البيانات. |
| يتم الحصول عليها من مدخلات المستخدمين أو من مدخلات محوسبة. | يتم الحصول عليها من عمليات معالجة البيانات. |

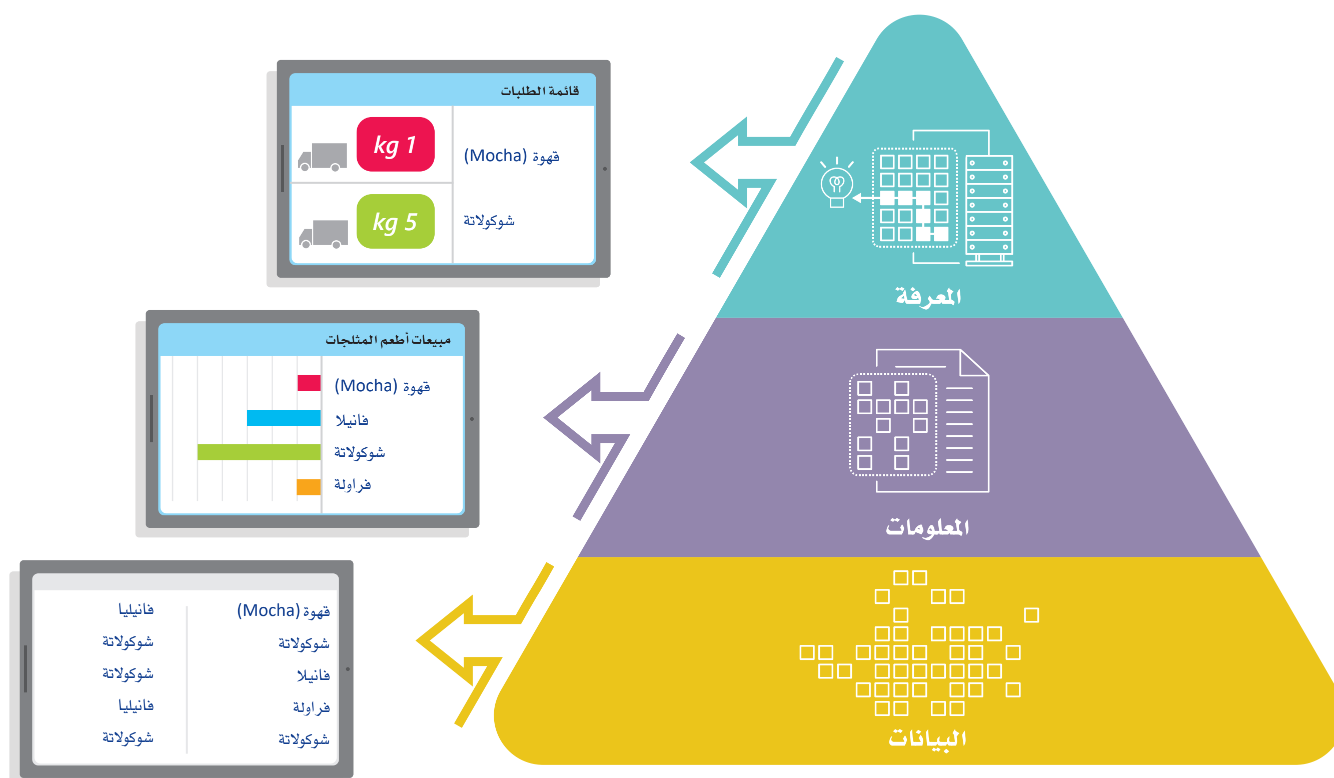
## المعرفة Knowledge

تعتبر المعرفة تمثيلًا لفهمك للعالم، وهي بشكل أساس مجموعة من المعلومات يتم استخدامها لتقديم فائدة أو تحقيق غرض معين. يمكنك القول أن فهم الشخص لبعض المعلومات حول شيء ما يوفر لديه معرفة به، فتصبح المعلومات معرفة عند تطبيق عمليات التفكير النقدي، أو التقييم، أو التخطيط، أو التنظيم.

لتشاهد المثال في الشكل 1.3، يمكنك ملاحظة أن البيانات الموجودة أسفل المخطط هي قائمة من الكلمات التي تخلو من أي سياق، وإذا تم تنظيم هذه البيانات وترتيبها، فيُمكن توفير بعض المعلومات. وعلى فرض أن هذه القائمة تحتوي على نتائج مبيعات نكهات المثلجات في اليوم السابق، يمكنك من خلال إجراء بعض عمليات التحليل عليها أن تحصل على بعض المعلومات المفيدة، فمثلًا يمكنك أن تدرك أن نكهة الشوكولاتة هي الأكثر مبيعًا.

تُمكّنُ المعرفة هنا مدير المتجر من اكتشاف أن نكهة الشوكولاتة هي الأكثر رواجًا، وهكذا يمكنه أن يطلب خمسة أضعاف كمية المثلجات بنكهة الشوكولاتة في المرة القادمة مقارنة بالنكهات الأخرى مثل نكهة القهوة (Mocha).






شكل 1.3: هرم البيانات – المعلومات – المعرفة

**جدول 1.2: أوجه الاختلاف بين المعلومات والمعرفة**

| | المعلومات | المعرفة |
|---|---|---|
| **المعنى** | بيانات تمت معالجتها لتصبح ذات سياق مفهوم. | استنتاجات مستقاة من المعلومات تساعد في اتخاذ القرارات. |
| **القدرة على التنبؤ** | لا تكفي وحدها للتوصل إلى استنتاجات أو قرارات. | تساهم في القدرة على التنبؤ واتخاذ القرارات. |
| **إيصال النتائج للآخرين** | يمكن إيصالها بسهولة من خلال الطرق الشفوية أو الورقية أو الإلكترونية. | تتطلب اطِّلاع بالموضوع المحدد. |
| **النتائج** | ينتج عنها الفهم. | ينتج عنها الإدراك. |
| **الهدف** | تجيب على أسئلة مثل مَن ومتى وماذا وأين. | تجيب على أسئلة مثل كيف ولماذا. |

## علم البيانات وذكاء الأعمال

## Data Science versus Business Intelligence

توجد البيانات في كل مكان من حولك، ويتم استخدامها ومعالجتها وتحليلها في جميع مجالات الحياة. تتطور نوعية البيانات واستخداماتها باستمرار، وتُستخدم بشكلٍ خاص في العديد من التطبيقات المهمة مثل **ذكاء الأعمال** (Business Intelligence)، ولهذا يُعتبر ذكاء الأعمال عملية قائمة على التقنية لتحليل البيانات وتوفير معلومات مهمة تساعد المدراء التنفيذين وغيرهم من المسؤولين وصنّاع القرار على اتخاذ قرارات دقيقة خاصة بالأعمال. وعلى الرغم من أن كلاً من علم البيانات وذكاء الأعمال يتضمن العمل على البيانات، إلا أنهما يختلفان عن بعضهما.

**ذكاء الأعمال (Business Intelligence):**

هو نظام مبني على البيانات ويشمل جمع وتخزين وتحليل وتمثيل البيانات لدعم عمليات اتخاذ القرارات.

يُعدُّ علم البيانات أكثر تعقيدًا مقارنة بذكاء الأعمال، حيث يقتصر نطاق ذكاء الأعمال على مجال الأعمال، ويتم فيه تحليل البيانات السابقة من خلال تطوير لوحات المعلومات وعرض مستخلصات (رؤى) الأعمال، وكذلك ترتيب و تنظيم وتحليل البيانات وذلك لاستخراج المعلومات التي من شأنها مساعدة الشركات على النمو وتحقيق أهدافها بناء على فهم الاتجاهات الحالية للأعمال. ويعتمد علم البيانات على استخدام البيانات المتوفرة للقيام بتنبؤات مستقبلية وعرض توقعات نمو الأعمال التجارية، وذلك بتوظيف مجموعة واسعة مما يسمى بالنماذج التنبؤية والخوارزميات الإحصائية المعقدة.

يتمثل الدور الأساسي لأدوات ذكاء الأعمال في تحليل معلومات المؤسسات والشركات والمساهمة في إعداد استراتيجيات الأعمال، أما أدوات عالم البيانات فتشمل أدوات معالجة البيانات وأدوات البيانات الضخمة وكذلك نماذج خوارزمية معقدة لتحليل البيانات واستخلاص التوصيات.

**جدول 1.3: أوجه الاختلاف بين علم البيانات وذكاء الأعمال**

| | علم البيانات | ذكاء الأعمال |
|---|---|---|
| **المدى** | تُستخدم البيانات لعمل تنبؤات مستقبلية لتطوير الأعمال. | تُحلل البيانات السابقة لاستنتاج الاتجاهات الحالية للأعمال. |
| **الأدوات** | يتضمن نماذج حسابية معقدة ومعالجة البيانات و أدوات البيانات الضخمة. | تقتصر الأدوات على تحليل المعلومات الإدارية والإشراف على استراتيجيات الأعمال. |
| **أنواع البيانات** | تتعامل بشكل أساسي مع البيانات غير المنظّمة وشبه المنظّمة، ويمكنها كذلك التعامل مع البيانات المنظمة. | تتعامل مع البيانات المنظّمة التي يتم تخزينها عادةً في مستودعات البيانات. |
| **التعقيد** | أكثر تعقيدًا مقارنة بذكاء الأعمال. | أبسط بكثير مقارنة بعلم البيانات. |
| **المرونة** | أكثر مرونة حيث يمكن إضافة مصادر البيانات حسب الحاجة. | أقل مرونة حيث يجب تصميم مصادر البيانات مسبقًا. |

## علم البيانات والذكاء الاصطناعي
## Data Science and Artificial Intelligence

كما تعرفت سابقًا على مفهوم علم البيانات، فإن مجال **الذكاء الاصطناعي** (Artificial Intelligence) يُعدُّ مجالًا آخر يتعامل مع كم كبير من البيانات.

يمكن استخدام كل تقنية منٍ هاتين التقنيتين بصورة منفصلة عن الأخرى للوصول لحلول لتحديات مختلفة، كذلك يمكن لكلٍ منهما إكمال بعضهما والتقارب معًا.

يختص علم البيانات بمعالجة البيانات التاريخية باستخدام أدوات حسابية للقيام بما يسمى **بالتحليل الوصفي** للبيانات (Descriptive Analysis) والذي يقدم وصفًا للمواقف المحددة، وكذلك للتنبؤ بالنتائج من خلال **التحليل التنبؤي** (Predictive Analysis)، ولتقديم الحلول والتوصيات للمشكلات من خلال **التحليل التوجيهي** (Prescriptive Analysis). من أكثر الأدوات استخدامًا هي الأدوات الإحصائية والإدارية التي يمكن بواسطتها تحليل البيانات المؤرخة. ومن ناحية أخرى يستخدم الذكاء الاصطناعي مجموعة متنوعة من التقنيات لمحاكاة الطريقة التي يفكر بها البشر والتي يقومون بناء عليها باتخاذ القرارات وتحليلها، فبدلًا من التركيز على إجراء الحسابات الرياضية، يتم التركيز عند استخدام أدوات الذكاء الاصطناعي على عناصر المعرفة والذكاء كعناصر حاسمة لحل المشكلات. ويهتم الذكاء الاصطناعي كذلك **بالحوسبة المعرفية** (Cognitive Computing). ومن المهم الإشارة إلى أن الفروقات التي ذُكرت سابقًا بين علم البيانات والذكاء الاصطناعي تبدو أقل وضوحًا في الاستخدامات العملية لهما لأن مشروعات علم البيانات المعقدة غالبًا ما تتضمن استخدام تقنيات تعلّم الآلة – أحد فروع الذكاء الاصطناعي – لتسهيل تحليل البيانات التنبؤي والتوجيهي.

يقدم علم البيانات والتعلم الآلي مساهمات كبيرة للعديد من المؤسسات عند استخدامهما بصورة منفصلة، إلا أن تقنيات تحليل البيانات التقليدية لا تتناسب مع العمل ببيانات غير كاملة أو غير دقيقة، أو في حال كانت السياقات التجارية أو العلمية تتغير بسرعة كبيرة مما يجعل البيانات تتقادم وتفقد قيمتها بسرعة كبيرة. كما تجدر الإشارة إلى أن تقنيات التعلم الآلي تتطلب قدرًا كبيرًا من البيانات نسبيًا.

يستخدم الجيل القادم من أدوات علم البيانات ومنصات ذكاء الأعمال تعلّم الآلة للقيام ببعض الإجراءات مثل التعرّف على الأنماط في البيانات لاكتشاف الأنماط المخفية وتقديم التصورات والرؤى المهمة لاتخاذ القرارات، ويُزوّد تعلّم الآلة والتعلّم العميق علم البيانات بتنبؤات أكثر دقة. إن توافر مجموعات البيانات الضخمة وانخفاض تكلفة معالجتها سحابيًا يُمكّن تعلّم الآلة من توفير إمكانات لم تكن ممكنة في الماضي. وعند الجمع بين علم البيانات والذكاء الاصطناعي، يمكن الحصول على طريقة فعالة جدًا في الحصول على نتائج دقيقة بشكل ملحوظ تساهم في اتخاذ قرارات أفضل وأسرع.

**الذكاء الاصطناعي**
**(Artificial Intelligence):**

أحد مجالات علوم الحاسب ويهدف لبناء أنظمة قادرة على أداء المهام التي تتطلب عادةً ذكاءً بشريًا مثل القدرة على التعلم والاستدلال، وحل المشكلات ومعالجة اللغة الطبيعية والإدراك.

### مثال

**أنشأت شركة أرامكو السعودية قسمًا جديدًا بمسمى قسم المصنع الرقمي للشركات (Corporate Digital Factory Department)، وتم تدعيم هذا القسم بعلماء البيانات وخبراء تعلّم الآلة للبحث عن التحديات التشغيلية وتطوير حلول ذكية للمساعدة في تحسين أداء الأعمال. تعمل الشركة بجد في الترويج للحلول المستوحاة من الذكاء الاصطناعي وذلك باستخدام المليارات من نقاط البيانات التي قام بجمعها الخبراء والجيولوجيون ومهندسو البترول على مدى عدة عقود.**

**ونظرًا لأن أرامكو كانت على الدوام من رواد استخدام تقنيات الذكاء الاصطناعي، تقوم الشركة باستخدام علم البيانات وأدوات التعلّم الآلي لتحسين عمليات استخراج البترول من المخزون الموجود تحت سطح البحر أيضًا، وتعمل تقنيات الذكاء الاصطناعي المتقدمة على تحسين خطط تطوير الحقول ومسارات الآبار؛ مما يؤدي إلى خفض التكلفة والحفاظ على البيئة. يستخدم علماء الجيولوجيا بالشركة أدوات الذكاء الاصطناعي في دراسة البيانات بشكل أسرع وأكثر كفاءة من أي وقت مضى. تعمل هذه العملية على تحسين فهم الخصائص البتروفيزيائية للتضاريس المراد استكشافها وحفرها وبالتالي تعزيز عملية اتخاذ القرارات.**

## دورة حياة علم البيانات
## Data Science Life Cycle

يقوم علماء البيانات والمتخصصين في العمل على مشروعات علم البيانات بتوظيف خبراتهم من خلال خطوات محددة لتنفيذ كل مشروع جديد بكفاءة.

يُطلق على هذه العملية اسم دورة حياة علم البيانات، وتتضمن خمس مراحل. تتميز كلٌ من المراحل المختلفة لهذه الدورة بخصائص معينة، مما يجعل من الممكن أن تشمل مشروعات خاصة مثل مشروعات الذكاء الاصطناعي وتعلّم الآلة، أو تمثيل العمليات الداخلية لمؤسسات معينة.




شكل 1.4: مراحل دورة حياة علم البيانات

### 1. تعريف المشكلة وصياغتها Problem Definition and Formulation

من أجل تصميم وإيجاد حل لمشكلة بواسطة علم البيانات، فإنك تحتاج أولًا إلى فهم ماهية المشكلة نفسها. يُعدُّ التحليل الشامل للمشكلة وبيئتها والمتغيرات التي تؤثر عليها أمرًا ضروريًا لتطوير الحلول اللازمة لتلك المشكلة، ويمكن أن يؤدي فهمك لمشكلة ما إلى تحسين إمكانية حلها أو إعاقة حلها بشكل كبير، وذلك لارتباطه المباشر بالنهج الذي سيتم اتباعه لحل تلك المشكلة. ويكمن الهدف التالي في تحديد الغاية المرجوة من هذا الحل، حيث أن مجموعة البيانات تتضمن دائمًا البيانات نفسها، ولكن طبيعة الإجابات التي تريد الوصول إليها قد تختلف حسب المشكلة المراد حلها.

**تعريف المشكلة وصياغتها:**

فهم أهداف ومتطلبات العمل أو المشكلة العلمية وتحويل هذه المعرفة إلى مسألة يمكن حلها بتحليل البيانات.

جدول 1.4: **أشهر أنواع تحليلات البيانات**

| | |
|---|---|
| **تحليل الانحدار** (Regression Analysis) | الحصول على الكميات أو الصفات الموجودة في مجموعة البيانات. |
| **تحليل التصنيف** (Classification Analysis) | تنظيم البيانات في فئات. |
| **التحليل العنقودي** (Clustering) | تنظيم البيانات في مجموعات. |
| **تحليل انحراف البيانات** (Anomaly Detection Analysis) | البحث عن انحراف أو شذوذ في البيانات. |
| **نظم التوصية** (Recommendation engines) | إعطاء توصية مستنيرة لمسألة محددة. |

## 2. جمع البيانات Data Collection

بعد أن يتم تحديد الأهداف، يجب توفير مجموعة البيانات نفسها، ورغم أنه قد يتم إدخال البيانات يدويًا أحيانًا، فمن المهم التنقيب وجمع البيانات، حيث يتعين في هذه المرحلة جمع بيانات كافية لمواصلة معالجتها. ويمكن أن تأتي البيانات نفسها من مجموعة متنوعة من المصادر، فمثلًا تقوم أجهزة الاستشعار البيئية وتطبيقات الهاتف المحمول ومنصات الويب بتوليد البيانات بصورة مستمرة ليتم تخزينها تلقائيًا في قواعد البيانات.

**جمع البيانات:**
عملية جمع القراءات أو الحقائق وتنسيقها، وتشمل الحصول عليها وتسميتها وتحسينها.

**جدول 1.5: تنسيقات تخزين البيانات الأكثر شيوعًا**

| | |
|---|---|
| الملفات المُنسقة (Formatted Files) | JSON و XML و CSV وجدول بيانات XLS. |
| قواعد البيانات العلائقية (Relational Databases) | خادم مايكروسوفت SQL وقاعدة بيانات أوراكل وأوراكل MySQL. |
| قواعد البيانات غير العلائقية (Non-Relational NoSQL Databases) | MongoDB و Azure Cosmos DB و AWS DynamoDB. |
| قواعد البيانات الرسومية (Graph Databases) | Neo4j و AWS Neptune و Dgraph. |
| قواعد بيانات السلاسل الزمنية (Time-series Databases) | InfluxDB وAWS Timescale. |

## 3. تجهيز البيانات وتنظيفها Data Preparation and Cleaning

تُعدُّ عملية "تنظيف" البيانات ومعالجتها أحد أهم المراحل في دورة حياة علم البيانات. يجب على عالم البيانات تصحيح وتجهيز البيانات التي تم جمعها في مرحلة التنقيب للتأكد من مناسبتها لمرحلة التحليل اللاحقة، وعند دمج البيانات من مصادر متعددة تزيد احتمالية تكرار البيانات أو تداخلها، الأمر الذي يتطلب عملية تصحيح وتصويب لتلك البيانات. وكذلك هو الحال إذا وُجدت بيانات تالفة أو منسقة بشكل غير صحيح أو مكررة أو خاطئة أو حتى غير مكتملة. تكمن أهمية تصحيح تلك البيانات في أن الرؤى أو الاستنتاجات المستمدة في مرحلة التحليل من تلك البيانات ستكون خاطئة وسيصعُب للغاية استنتاج ما إذا كانت المشكلة ناشئة من أخطاء في خطوات التحليل أو أن البيانات نفسها لم يتم تصحيحها، ولهذا السبب فإن عملية تنظيف البيانات والتحقق من صحتها جيدًا قبل تحليلها تُعدُّ أمرًا مهمًا للغاية للعملية بأكملها.

**تنظيف البيانات:**
عملية متعددة المراحل لمراجعة البيانات وتصحيحها للتأكد من أنها في صيغة موحدة، ويتضمن ذلك معالجة القيم المفقودة والبيانات المشوشة، وحل التناقضات والتكرارات.



## 4. التحليل الاستكشافي للبيانات Exploratory Data Analysis

**التحليل الاستكشافي للبيانات:**

هو نهج لتحليل مجموعات البيانات لتلخيص خصائصها الرئيسة، ويتم عادة باستخدام الأساليب المرئية.

بعد أن جمعت البيانات وقمت بتصحيحها، يمكنك تحليل مجموعة البيانات واستنباط الإجابات المطلوبة لأسئلتك، ويتم إجراء تحليل البيانات باستخدام أدوات تحليل البيانات أو الأكواد والمكتبات البرمجية المتخصصة، وقد يكون التحليل بسيطًا وذلك بدراسة متغير واحد أو أكثر، وقد يتسع ليشمل عمليات أكثر تعقيدًا تتضمن عمليات إحصائية متقدمة.

يُعدُّ تعلّم الآلة من أكثر الطرق شيوعًا في الوقت الحالي لتحليل مجموعة البيانات، ويجب اتّباع خطوات محددة لتحليل البيانات باستخدام تعلّم الآلة، ففي البداية يجب تحديد نموذج تعلّم الآلة بإيجاد قيم المدخلات والمخرجات يليها بناء خوارزمية التحليل نفسها.

تعتبر هذه العملية معقدة، ولهذا فإن هناك متخصصين للقيام بها مثل علماء البيانات ومهندسي تعلّم الآلة. بعد الانتهاء من الخوارزمية، يتم تجريب النموذج واختباره، وعند اكتمال هاتين المرحلتين يمكنك استخدام البيانات الناتجة منه للوصول للإجابات المرجو الحصول عليها من عمليات التحليل.

## 5. التمثيل الرسومي للبيانات Data Visualization

**تمثيل البيانات:**

يسلط التمثيل الرسومي للمعلومات الضوء على أنماط واتجاهات البيانات، ويساعد القارئ على تطوير رؤى وتوصيات بناءً على تلك البيانات.

يتم تقديم البيانات التي يتم تحليلها عادة بصورة جداول بيانات، مما يتيح لمحللي البيانات ذوي الخبرة استخدامها، ويقدم التمثيل المرئي لتحليل البيانات إمكانية استخلاص رؤى وتوصيات ذات جودة أفضل، بينما توفر الرسوم البيانية والمخططات وحتى الخرائط، وكذلك التقارير المنسّقة طريقة فعالة لرؤية وفهم أنماط البيانات واتجاهاتها أي ما توحي به تلك البيانات.

يُعدُّ تمثيل النتائج أمرًا ضروريًا لاتخاذ قرارات مُستندة إلى البيانات عند التعامل مع كميات هائلة من المعلومات.




الشكل 1.5: تحليل تفشي فيروس كورونا (COVID-19) باستخدام التحليلات المرئية. معهد ساس– SAS. © 2022 SAS Inc.



# تمرينات

**1**

| حدد الجملة الصحيحة والجملة الخاطئة فيما يلي: | صحيحة | خاطئة |
|---|---|---|
| 1. علم البيانات هو حقل متعدد التخصصات يركّز على استخراج معلومات ذات فائدة من البيانات. | ⚪ | ⚪ |
| 2. يُطلق اسم المعرفة على البيانات عند تحليلها وتنظيمها وهيكلتها لتصبح ذات معنى. | ⚪ | ⚪ |
| 3. يتم الحصول على المعلومات من خلال عمليات تحليل البيانات. | ⚪ | ⚪ |
| 4. يطلق اسم المعرفة على عملية جمع البيانات بطريقة صحيحة تجعلها ذات فائدة. | ⚪ | ⚪ |
| 5. تُعدُّ الرسوم البيانية والمخططات من وسائل عرض المعلومات. | ⚪ | ⚪ |
| 6. تعتبر معلومات حالة الطقس المقدمة من الأرصاد الجوية بمثابة معرفة. | ⚪ | ⚪ |
| 7. إن علم البيانات، والذكاء الاصطناعي، وذكاء الأعمال هي ثلاث مجالات مختلفة ومستقلة عن بعضها البعض. | ⚪ | ⚪ |
| 8. يساعد استخدام التمثيل المرئي لتحليل البيانات على استنباط رؤى أفضل مما يعني اكتساب معرفة أفضل بمعنى تلك البيانات. | ⚪ | ⚪ |
| 9. تُعدُّ نُظم التوصية الذكية وتحليل الانحدار من أفضل طرق تخزين البيانات. | ⚪ | ⚪ |
| 10. تُعدُّ قواعد بيانات السلاسل الزمنية وقواعد البيانات غير العلائقية (NoSQL) جزءًا من وسائل جمع البيانات. | ⚪ | ⚪ |



**2** أنشئ قائمة من البيانات ثم حول تلك البيانات إلى معلومات مفيدة، ثم وضّح كيف يحوّل الحاسب البيانات إلى معلومات؟

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

**3** وضّح الفروق الثلاثة الرئيسة بين علم البيانات والذكاء الاصطناعي، وادعم إجاباتك ببعض الأمثلة.

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

**4** وضّح وقارن بين علم البيانات وذكاء الأعمال، وإذا كان لديك شركة استثمارية، أي الحقلين سيكون خيارك المفضل للاستثمار؟

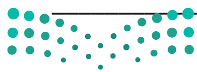
________________________________________

________________________________________

________________________________________

________________________________________

________________________________________



**5** ما مدى فعالية تقارب علم البيانات والذكاء الاصطناعي؟ ابحث في الإنترنت واذكر مثالين ناجحين على ذلك.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**6** وضِّح المقصود بعلم البيانات، واذكر ثلاثة تطبيقات حياتية في المجال الصحي، ومجال الأعمال التجارية، والترفيه، ثمّ بيّن لماذا يُعدُّ علم البيانات ضروريًا لمثل هذه المجالات؟

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

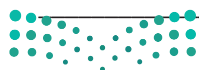
**7** وضِّح وقارن بين مجموعات البيانات المعالجة وغير المعالجة التي تَصِفُ الدرجات الفصلية للطالب وأدائِه خلال العام الدراسي.

ما الانطباعات والأفكار التي يمكنك الحصول عليها من مجموعات البيانات هذه؟ وهل يمكنك توقّع الأداء الأكاديمي المستقبلي للطالب في الجامعة من خلال هذه البيانات؟

______________________________________________

______________________________________________

______________________________________________



**8** ابحث عن مزيد من المعلومات حول "مصنع أرامكو السعودية الرقمي" وحدد ثلاثة أمثلة لاستخدام الذكاء الاصطناعي في استخراج البيانات، ومن ثم اعطينا رأيك في مدى تأثير هذه التقنيات على العمليات التشغيلية للشركة؟

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**9** ابحث في الإنترنت بصورة مفصّلة عن نماذج دورة حياة علم البيانات التي تصف المراحل الرئيسة المذكورة في هذا الدرس، ومن ثم اختر إحداها وحدد المراحل الإضافية واشرحها بإيجاز.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________



رابط الدرس الرقمي



www.ien.edu.sa

الدرس الثاني

# التعامل مع البيانات

## ما المقصود بالبيانات الضخمة؟ What is Big Data?

يشير مصطلح **البيانات الضخمة** (Big Data) إلى البيانات الكبيرة جدًا أو المعقدة التي لا يمكن معالجتها بالطرق التقليدية، ونظرًا لأن كم هذه البيانات يُعدُّ كبيرًا جدًا لتتمّ معالجتها باستخدام أنظمة الحوسبة التقليدية، فإن تخزين مجموعاتها ومعالجتها يعتبر تحديًا كبيرًا، وكذلك قد تتطلب السرعة الهائلة لعملية جمع البيانات متطلبات تخزين عالية للغاية.

**البيانات الضخمة (Big Data):**

مجموعة بيانات كبيرة تتطلب تقنيات قابلة للتوسع لتخزينها ومعالجتها وإدارتها وتحليلها وذلك نظرًا لخصائص حجمها، وتنوعها وسرعتها وتباينها وبالطبع قيمتها.

## خصائص البيانات الضخمة Characteristics of Big Data

هناك خمسة معايير أساسية تساعدنا في تصنيف أي بيانات تحت مصطلح "البيانات الضخمة" وهي: التنوع، والقيمة، والحجم، والموثوقية، والسرعة. وتعتبر البيانات "ضخمة" عندما تأتي بأحجام كبيرة، وبمعدل سريع جدًا، وبتنوع كبير، وبدقة عالية ،وفائدة. ويجب أن تستوفي البيانات جميع هذه المعايير لكي يتم اعتبارها "بيانات ضخمة".

### التنوع Variety

يشير التنوع إلى العديد من أنواع البيانات المتوافرة،ويتم هيكلة البيانات التقليدية المختلفة وتكييفها بدقة في قواعد البيانات العلائقية، ولكن مع ظهور البيانات الضخمة، أصبحت البيانات تتوافر في أنواع جديدة غير منظّمة. تتطلب أنواع البيانات غير المنظمة وشبه المنظمة (مثل النصوص والصوت والفيديو) معالجة إضافية مسبقة لاستخلاص المعاني ودعم معلومات البيانات الوصفية المتعلقة بتلك البيانات، وبدون هذه البيانات الوصفية يكون من المستحيل معرفة ما يتم تخزينه وكيف يمكن معالجته.




الشكل 1.6: خصائص البيانات الضخمة – المعايير الخمسة

### القيمة Value

إن جمع الكثير من البيانات لا يعني أن تلك البيانات هي ذات قيمة، فقيمة البيانات تتمثل في إمكانية الحصول على التوصيات والوصول إلى بعض الأفكار من خلالها. يشير مصطلح القيمة إلى مدى فائدة البيانات في اتخاذ القرارات، وبالطبع فإن إجراء التحليلات المناسبة هو وسيلة استخراج قيمة البيانات الضخمة.

### الحجم Volume

نظرًا لأنه يجب معالجة كميات كبيرة من البيانات غير المنظمة والتي تتميز بقلة الكثافة وتسمى (low density data)، فإن كم البيانات يعد جانبًا مهمًا في البيانات الضخمة. يمكن أن تكون قيمة بعض هذه البيانات غير معروفة قبل القيام بتحليلها، مثل بيانات تصفح المستخدمين لأحد مواقع الويب أو أحد تطبيقات الهاتف الذكي، أو تلك البيانات التي يتم الحصول عليها من أجهزة إنترنت الأشياء المدعمة بأجهزة الاستشعار. قد يصل حجم هذه البيانات إلى العشرات، بل المئات من التيرابايت من البيانات.

### الموثوقية Veracity

ترتبط صحة البيانات بمدى دقة مجموعة البيانات أو موثوقيتها. لا ترتبط الموثوقية بجودة البيانات نفسها فحسب، بل أيضًا بمدى مصداقية مصدر البيانات ونوعها وكيفية معالجتها.

### السرعة Velocity

يشير مصطلح السرعة إلى معدل التقاط البيانات وتخزينها. تنتج البيانات من معظم الأجهزة الذكية المتصلة بالإنترنت (أجهزة إنترنت الأشياء) والأجهزة المحمولة في الوقت الحقيقي أو قريبًا من الوقت الحقيقي، مما يتطلب الجمع الفوري لتلك البيانات وكذلك نقلها وتخزينها.

## تقنيات إدارة البيانات الضخمة
## Technologies that Enable the Management of Big Data

تستخدم الشركات أنظمة الحاسب وقواعد البيانات للاحتفاظ بالسجلات المختلفة مثل المعاملات المتعلقة بمعالجة الطلبات والمدفوعات وتتبع العملاء وإدارة التكلفة في الشركات. تحتاج الشركات أيضًا إلى نظام لإعداد التقارير لتوفير المعلومات التي تساعدها على العمل بكفاءة ولمساعدة المدراء التنفيذيين على اتخاذ القرارات المدروسة التي تضمن أداء أفضل للأعمال.

يحتاج مديرو المتجر الإلكتروني إلى تحسين تجربة الشراء والتأكد من أن زوار الموقع الذين يتصفحون المنتجات سيصبحون زبائن للمتجر وذلك من خلال شراء المنتجات، وكذلك العمل على عودة الزبائن للشراء مرات أخرى في المستقبل من خلال الموقع. يمكن للشركة تحليل جميع البيانات التي يتم جمعها أثناء تصفح الزوار للمتجر الإلكتروني على الويب أو من خلال تطبيق الهاتف الذكي، وتتضمن تلك البيانات تفاصيل دقيقة عن تصفح الزوار للموقع، بما فيها أماكن وضع المؤشر على الشاشة وأجزاء الموقع التي يقضون وقتًا أطول في تصفحها، ومدة المرور فوق المنتج قبل الضغط للحصول على المزيد من المعلومات عنه أو للقيام بشرائه بالفعل. ينتج عن هذه التفاصيل الدقيقة التي يتم جمعها كم هائل من البيانات التي يجب تحليلها لتقديم رؤية واضحة وقيمة للقائمين علي أعمال الشركة. يتم استخدام نتائج تحليل تلك المعلومات لإحداث تغييرات في مخطط موقع الويب أو المتجر، ولتعديل أسعار المنتجات سواء بالزيادة أو بالخصم، ولتنظيم الحملات التسويقية للمنتجات على وسائل التواصل الاجتماعي للتأثير على سلوكيات الشراء لدى الزبائن.

يتطلب القيام بهذا الأمر من الشركات توفير تقنيات وأدوات جديدة لإدارة وتحليل البيانات الضخمة لاستخراج قيمة الأعمال، ويجب جمع البيانات المطلوبة من المصادر الداخلية كدوائر المبيعات والتصنيع والمحاسبة، وكذلك من المصادر الخارجية كالبيانات الإحصائية عن النمو السكاني وطبيعة الزبائن وأعمارهم، وكذلك البيانات المتعلقة بالشركات المنافسة مثلًا، وذلك لاستخراج معلومات موجزة وموثوقة حول الوضع الحالي والمستقبلي للشركة والتأثيرات المحتملة لمتغيرات السوق. تحتوي البنية التحتية الحديثة لذكاء الأعمال على مجموعة من الأدوات والتقنيات لتخزين البيانات ومعالجتها للحصول على معلومات مفيدة من البيانات الضخمة، وتشمل هذه التقنيات مستودعات البيانات وبحيرات البيانات وعمليات الحوسبة في الذاكرة.



## مستودعات البيانات Data Warehouse

قد تعتبر مستودعات البيانات الأداة الأقدم لتحليل بيانات الشركات. يشير مستودع البيانات إلى قاعدة البيانات التي تخزن البيانات الحالية والتاريخية التي نتجت عن العديد من أنظمة المعاملات التشغيلية الأساسية مثل أنظمة المبيعات، ودعم العملاء، والتصنيع، والتي تجعل البيانات متاحة لصانعي القرار في الشركة، ويتم دمج هذه البيانات مع البيانات من المصادر الخارجية لتحويل البيانات غير المكتملة إلى بيانات منظّمة قبل تخزينها في مستودع البيانات. يوفر نظام مستودع البيانات أيضًا مجموعة من الأدوات للتحليل والاستعلام وكذلك أدوات إعداد التقارير الرسومية.

## الحوسبة في الذاكرة In-Memory Computing

هي طريقة لتسهيل عملية تحليل البيانات الضخمة لاعتمادها بصورة أساسية على **ذاكرة الحاسب الرئيسة** (RAM) لتخزين البيانات. يصل المستخدمون إلى البيانات المخزنة في الذاكرة الأساسية للنظام وبالتالي يتم تجاوز معوقات استرداد وقراءة البيانات الموجودة في قاعدة البيانات التقليدية المستندة إلى التخزين على الأقراص مما يعني تقليل وقت الاستعلام بشكل كبير. تتميز الخوادم السحابية بشكل خاص بوجود سعة كبيرة من ذاكرة الوصول العشوائي، مما يسهل استخدامها في عمليات الحوسبة في الذاكرة.

## بحيرة البيانات Data Lake

بحيرة البيانات هي مستودع بيانات عادةً ما يكون سحابيًا يُستخدم لتخزين كميات هائلة من البيانات الأولية وغير المعالجة. في هذه الطريقة يتم استخدام عنوان URL ثابت لدعم كلٍ من البيانات المنظمة (مثل قواعد البيانات) والبيانات غير المنظمة (مثل رسائل البريد الإلكتروني والمستندات).

يعتبر التمييز بين هذه التقنيات الثلاثة مهمًا نظرًا للأغراض المختلفة لاستخدامها ولكيفية عملها في معالجة البيانات بشكل صحيح. لا تعمل هذه التقنيات معًا، ولكن يتم اختيار إحداها اعتمادًا على نوع الشركة، فقد تكون بحيرة البيانات هي الخيار الجيد لإحدى الشركات، بينما يعد مستودع البيانات الخيار الأفضل لشركة أخرى.

## التنقيب في البيانات الضخمة Mining Big Data

كما عرفت سابقًا، يتم جمع البيانات الضخمة باستمرار بواسطة أجهزة الاستشعار والتطبيقات العامة والتطبيقات الشخصية. إن عملية جمع البيانات ليست سوى الخطوة الأولى في العملية المشار إليها باسم اكتشاف المعرفة.

يشير اكتشاف المعرفة إلى العملية الشاملة للوصول إلى المعرفة المفيدة من البيانات، ويشير التنقيب في البيانات إلى خطوة معينة في هذه العملية، فالتنقيب عن البيانات هو تطبيق لخوارزميات محددة لاستخراج الأنماط من البيانات، وتحديد العلاقات المختلفة داخل هذه البيانات. تعتبر الخطوات الأخرى في عملية اكتشاف المعرفة مثل تنظيف البيانات، وتكامل البيانات، وتحويل صيغة البيانات، والتفسير الصحيح لنتائج التنقيب ضرورية لضمان اشتقاق المعرفة المفيدة من البيانات (انظر الجدول 1.6).

**التنقيب في البيانات (Data Mining):**
عملية اكتشاف الأنماط في كمية كبيرة من البيانات واستخراج المعلومات المفيدة في توقع السلوك المستقبلي.

| بعض المهام الرئيسة التي يتم إنجازها عن طريق التنقيب في البيانات: |
|---|
| تحليل البيانات لاكتشاف الأنماط والاتجاهات. |
| صياغة التنبؤات لمدخلات مجموعات البيانات المختلفة. |
| تصنيف أو تجميع أو توقع القيم المختلفة لمجموعة البيانات. |
| تسهيل عملية اتخاذ القرارات المدروسة. |



**الجدول 1.6: خطوات اكتشاف المعرفة**

| | |
|---|---|
| **تصحيح البيانات:** | تنظيف البيانات التالفة وغير المطابقة، وإزالة أنواع البيانات الخاطئة وما إلى ذلك. |
| **تكامل البيانات:** | يحدث التنقيب في البيانات من مصادر متعددة. يجب دمج مصادر البيانات هذه في مجموعة بيانات واحدة. |
| **اختيار البيانات:** | تحديد جزء مجموعة البيانات الذي يجب استخدامه لعملية استخراج البيانات. من المهم تحديد مجموعة البيانات الأكثر مواءمة لأهدافك لأن استخراج البيانات مهمة تستغرق وقتًا طويلًا. |
| **تحويل صيغة البيانات:** | يُعدُّ إعداد مجموعات البيانات الأولية وتنسيقها أمرًا ضروريًا لأن عمليات التنقيب عن البيانات تحتاج إلى أن يكون لمدخلاتها تنسيق محدد لتحليلها. |
| **التنقيب في البيانات:** | هي العملية الفعلية لتحليل البيانات واستخراج النتائج المرجوة من التحليل من خلال الأنماط. |
| **تقييم النمط:** | تقييم الأنماط التي تم إنشاؤها خلال خطوات التنقيب عن البيانات، وتحديد أيها مفيد لكل هدف محدد. |
| **تمثيل المعرفة:** | تمثيل النتائج التي تم الحصول عليها من خلال التقارير، والرسوم البيانية والمخططات الواضحة والمختصرة. |

## البيانات الضخمة والتخزين السحابي Big Data and Cloud Storage

هناك خياران معتمدان لتخزين البيانات الضخمة: التخزين السحابي والتخزين الداخلي، ولقد كان تطوير تطبيقات البيانات الضخمة في الماضي يعتمد أساسًا على حفظ البيانات في وسائط التخزين داخليًا (على الخوادم داخل الشركات والمؤسسات)، مما تطلب توفر مستودعات بيانات محلية عالية التكلفة، وكذلك تثبيت برامج معقدة لإدارة تلك المستودعات. ساهمت التطورات الحديثة في علوم الحوسبة والبيانات في استبدال تلك الطريقة بالتخزين السحابي، والذي يعد بمثابة الحل الأمثل لتخزين البيانات الضخمة، وذلك لما يلي:

أ) توافر النطاق العريض عالي السرعة على نطاق واسع يسهل حركة البيانات من مكان إلى آخر. ومع وجود بيانات منتجة محليا لم تعد هناك حاجة لتخزين البيانات داخليًا، بل أصبح بالإمكان نقلها إلى التخزين السحابي لتحليلها.

ب) أصبحت غالبية التطبيقات تعتمد على التخزين السحابي، مما يعني أن عملية إنتاج المزيد من البيانات وتخزينها سحابيًا تزداد باستمرار، ولقد ساهم ذلك في قيام أعداد متزايدة من رواد الأعمال بعمل تحليلات جديدة للبيانات الضخمة لمساعدة الشركات على تحليل البيانات السحابية في كثير من المجالات مثل معاملات التجارة الإلكترونية وبيانات أداء تطبيقات الويب.

هناك جوانب متعددة للتخزين السحابي تجعله خيارًا أفضل للشركات، فمثلًا يمكن أن يشمل التخزين السحابي أنظمة تخزين البيانات الضخمة وكذلك أنظمة النسخ الاحتياطي.

وتوجد الكثير من الخيارات المتاحة من قبل مزودي الخدمات مثل أمازون ومايكروسوفت وقوقل للتخزين السحابي مع توفيرها حماية البيانات والخصوصية، وبالطبع قابلية التوسع والتكلفة المعقولة لهذه الخدمات.



شكل 1.7: مركز بيانات يوفر خدمة التخزين السحابي

باستخدام النسخ الاحتياطي السحابي للبيانات الضخمة، يمكن للمؤسسات الاستفادة من الخدمات من مراكز البيانات التي تمتد عبر مواقع جغرافية متعددة، مما يضمن التوافر الدائم واستعادة البيانات بسهولة. وباستخدام التخزين السحابي يمكن نسخ البيانات احتياطيًا عبر عدة مراكز بيانات في مناطق مختلفة من العالم، وبذلك لا يتم الاحتفاظ بالنسخ الاحتياطية في مكان واحد.

توفر تقنيات التخزين السحابي خصائص أخرى للحماية للنسخ الاحتياطية حيث يضمن مقدمو تلك التقنيات حماية البيانات المنسوخة احتياطيًا إلى التخزين السحابي عبر تقنيات التشفير المتقدمة قبل القيام بنقل البيانات وخلال نقلها وكذلك بعد نقلها.

وكما ذُكر سابقًا، فإن معالجة البيانات الضخمة تتطلب سعة تخزين وقوة معالجة، أما من حيث السعة التخزينية، فإن التقنيات السحابية تفي بهذا الأمر، ويمكن للشركات الحصول على خدمات التخزين القابلة للتوسع بيسر. يمكن لهذه التقنيات أيضًا تلبية متطلبات الحوسبة لتحليل البيانات الضخمة، وقد أصبح خبراء تحليلات البيانات يوصون باستخدام الخدمات المدعومة سحابيًا للقيام بعمليات التحليل لإدراكهم بالإمكانيات اللانهائية التي توفرها التقنيات.

## مزايا وعيوب تخزين البيانات الضخمة سحابيًا Pros and Cons of Big Data Cloud Storage

يمكن من خلال الجمع بين تحليلات البيانات الضخمة والحوسبة السحابية إيجاد فرص لم تكن ممكنة من قبل لفهم البيانات والحصول على المعرفة ودعم عمليات اتخاذ القرارات. يجب عليك النظر إلى المزايا وكذلك إدراك التحديات التي تتعلق باستخدام الحوسبة السحابية في العمليات المتعلقة بالبيانات الضخمة.

**جدول 1.7: مزايا وعيوب تخزين البيانات الضخمة سحابيًا**

| المزايا | العيوب |
|---|---|
| تتطلب الكميات الكبيرة من البيانات المنظمة وغير المنظمة توفر شبكات ذات نطاق ارسال واسع وذلك لسرعة الإرسال والتخزين. يوفر التخزين السحابي بنية تحتية متاحة بسهولة مع القدرة على التوسع للتعامل مع أي مقدار من حركة مرور البيانات ومتطلبات التخزين. | تقدم إمكانيات تحكم مباشر أقل في أمن البيانات، وقد تتعرض لعمليات تؤدي إلى انتهاك البيانات، وبالتالي إلى عواقب خطيرة فيما يتعلق بلوائح خصوصية البيانات. |
| يؤدي تخزين البيانات الضخمة سحابيًا إلى التخلص من الحاجة إلى الاحتفاظ بأجهزة وبرامج وموظفين متخصصين عند الحاجة، ويُعدُّ نموذج الحوسبة السحابية المبني على الدفع حسب الحاجة إلى الخدمات أكثر فعالية من حيث التكلفة، مما يساهم في خفض التكلفة وزيادة الكفاءة والحد من هدر الموارد. | يمكن لمزود الخدمة السحابية رفع تكلفة الخدمات التي يقدمها في أي وقت، مما يعني ارتفاع التكلفة لأعمال الشركات المستخدمة لهذه الخدمات، والتي لا يمكنها الانتقال بسهولة إلى مقدم خدمات آخر يقدم أسعارًا تنافسية. |
| تركِّز الشركة على عمليات تحليل البيانات بدلًا من إدارة البنية التحتية، مما ينعكس بشكل إيجابي على الأداء والميزة التنافسية. | يعني تخزين البيانات الضخمة سحابيًا أن توفر البيانات يعتمد على الاتصال بالشبكة. تؤثر المشاكل المتعلقة بالشبكات كتدني جودة الاتصال أو **تأخُّر الاستجابة** (latency)، والتي قد تظهر في البيئة السحابية على سرعة جمع البيانات ومعالجتها وتخزينها. |



## سياسة الشركات وحوكمة البيانات
## Data Governance and Policies

تحدد الضوابط والهياكل التنظيمية للشركات والمؤسسات المسؤوليات وطرق اتخاذ القرارات المتعلقة بإدارة البيانات، والتي تتضمن تطوير السياسات والإجراءات الداخلية التي تتحكم بإدارة البيانات.

تساعد إدارة البيانات المؤسسات الخاصة أو المؤسسات الحكومية وغير الربحية في التعامل مع عمليات إدارة البيانات بجودة عالية خلال جميع مراحل دورة حياة البيانات، وتؤدي هذه السياسات والإجراءات الفعالة إلى تحسين الأعمال والنتائج، حيث تقوم الشركات والمؤسسات بجمع كميات هائلة من البيانات الداخلية والخارجية، وتعتبر إدارة البيانات ضرورية لاستخدام تلك البيانات بفعالية وإدارة المخاطر وخفض التكاليف المختلفة.

**تضمن حوكمة البيانات أن البيانات:**

| |
|---|
| آمنة. |
| موثوقة. |
| مُوثقة. |
| مُدارة. |
| مُدققة. |

### أهمية حوكمة البيانات The Importance of Data Governance

من غير الممكن تصحيح التناقضات الموجودة في بيانات الأنظمة المختلفة داخل المؤسسة بدون إدارة مناسبة للبيانات. فعلى سبيل المثال، قد يتم تخزين أسماء العملاء في أنظمة المبيعات وأنظمة خدمة العملاء بطرق مختلفة.

قد يُصعّب هذا الأمر من تكامل البيانات ويؤثر على دقة ذكاء الأعمال وعلى إعداد التقارير. كذلك فقد لا يتم اكتشاف أخطاء البيانات الموجودة وبالتالي لا يتم تصحيحها في الوقت المناسب، مما يعرض سلامة البيانات للخطر.

أصبح واجبًا على المؤسسات أن تمتثل للتشريعات الجديدة الخاصة بخصوصية البيانات وحمايتها مثل اللائحة العامة لحماية البيانات في الاتحاد الأوروبي (GDPR) وقانون خصوصية المستهلك في كاليفورنيا (CCPA)، وذلك لأن حوكمة البيانات بصورة سيئة تجر المؤسسات إلى صعوبات و تجعلها تحت طائلة مواجهة العقوبات. يُنظّم قانون حماية البيانات الشخصية الجديد في المملكة العربية السعودية (PDPL) موضوع معالجة البيانات الشخصية، ويعد بمثابة التشريع الأول الخاص بخصوصية البيانات الحكومية في المملكة العربية السعودية. يشمل هذا التشريع جميع الصناعات وأنواع المؤسسات، ويشرف مكتب إدارة البيانات الوطنية (NDMO) على اللوائح الجديدة ويفرضها، وينطبق هذا القانون أيضًا على المؤسسات الأجنبية العاملة في المملكة العربية السعودية التي تتعامل مع البيانات الشخصية للمواطنين والمقيمين في المملكة، وبشكل خاص تلك البيانات الصحية والائتمانية والمالية.

**تتطلب بعض أنواع البيانات كالبيانات المالية والصحية معالجة بشكل أكثر دقة من البيانات الأخرى، فعادةً ما يتم تنظيم البيانات الصحية بشكل جيد بدءا من مرحلة جمع تلك البيانات حتى إعداد التقارير ونشر المعلومات. يتفهم جميع من يهمهم الأمر تمامًا المخاطر المتعلقة بالخصوصية وتلك القيود التي تفرضها التشريعات، وبالتالي فإن تحديد إطار عمل حوكمة البيانات في المؤسسات الصحية كالمستشفيات يُعد أمرًا مهمًا للغاية.**

### مكونات إطار عمل حوكمة البيانات
### Data Governance Framework Components

تشكل السياسات والضوابط التوجيهية والعمليات والهياكل التنظيمية والتقنيات المُنفذة كجزء من برنامج الحوكمة إطار عمل لحوكمة البيانات، ويحدد إطار العمل أيضًا مهمة البرنامج وأهدافه والمسؤوليات المختلفة وكيفية قياس النجاح في المهام التي سيتم تضمينها في البرنامج.

يجب إنشاء إطار الحوكمة في أي مؤسسة ونشره داخليًا وشرح آلية عمله بحيث يكون لدى كل مشارك فهم واضح منذ البداية.



## معايير حوكمة البيانات Data Governance Standards

قامت منظمة المعايير الدولية ISO بتطوير معيار ISO/IEC 38505 لتطبيق مبادئ حوكمة تقنية المعلومات على متطلبات إدارة البيانات.

**جدول 1.8: المبادئ الستة لحوكمة البيانات**

| | |
|---|---|
| **المسئولية** | تُعيّن للأفراد. |
| **الاستراتيجية** | تتوافق مع مهمة ورؤية المؤسسة. |
| **الحيازة** | تتوافق مع المتطلبات التنظيمية. |
| **التوافق** | ضمان الامتثال للتشريعات والسياسات الداخلية وأخلاقيات العمل. |
| **الأداء** | تلبية متطلبات المؤسسة. |
| **السلوك الإنساني** | تشجيع الناس على المشاركة. |

## معايير إدارة البيانات السعودية
## Saudi Data Management Standards

على غرار متطلبات إدارة البيانات الدولية ISO/IEC 38505، طوّر مكتب إدارة البيانات الوطنية (NDMO) ضوابط إدارة البيانات الوطنية وحماية البيانات الشخصية في المملكة. يعتبر مكتب إدارة البيانات الوطنية مسؤولًا عن تنفيذ الضوابط والسياسات وآليات الحوكمة والضوابط الخاصة بالبيانات والذكاء الاصطناعي ومتابعة الامتثال من قبل المنظمات والشركات. تنطبق المعايير على جميع البيانات بغض النظر عن النموذج أو النوع بما فيها السجلات الورقية والبيانات الرقمية والتسجيلات الصوتية و الصور ومقاطع الفيديو، وكذلك المستندات المكتوبة بخط اليد أو أي شكل من أشكال البيانات المسجلة.



https://sdaia.gov.sa/ndmo




شكل 1.8: نماذج لصفحات من مكتب إدارة البيانات الوطنية – ضوابط ومواصفات إدارة البيانات الوطنية وحوكمتها وحماية البيانات الشخصية © الهيئة السعودية للبيانات والذكاء الاصطناعي

### مثال

أنشأت شركة اتصالات سعودية قطاع تحليل البيانات المؤسسية (CAD) عام 2016 للمساعدة في تحقيق هدفها المتمثل في تقديم أفضل ممارسات حوكمة وإدارة البيانات. تعتبر البيانات والأشخاص والعمليات والتقنيات من أهم ركائز الشركة في الحوكمة، حيث تتضمن المبادرة تحقيق جميع الركائز لأجل تحول رقمي ناجح.

تسعى الشركة الآن إلى اعتماد حلول حوكمة البيانات المبتكرة التي تعزز الذكاء الاصطناعي وتوسع مفهوم حوكمة البيانات إلى "حوكمة التحليلات".

يهدف هذا السعي إلى تحقيق تغيير إيجابي في الأعمال وذلك من خلال تحديد المتطلبات وخطوات سير الأعمال بشكل واضح.

## حوكمة البيانات وإدارتها

### Data Governance versus Data Management

من الأهمية بمكان إدراك أن حوكمة البيانات هي أحد مكونات إدارة البيانات الشاملة. إن وضع القواعد الإرشادية لحوكمة البيانات دون التنفيذ الفعلي لها يعتبر مضيعة للوقت والجهد دون معنى أو قيمة حقيقية، فحوكمة البيانات تحدد جميع الضوابط والسياسات والعمليات، والتي تُنفذ بواسطة إدارة البيانات، والتي مهمتها هي جمع البيانات واستخدامها في صنع القرار من خلال اتباع أساسيات الحوكمة والتي تتمثل بالضوابط والسياسات والعمليات المتعلقة بالبيانات. تشبه حوكمة البيانات عملية تطوير التصميم لبناء منزل جديد، أما إدارة البيانات فهي عملية البناء نفسها، فإذا تخيلت هذا الأمر، فيمكنك بناء منزل دون تصميم، ولكن هذا البناء سيكون أقل كفاءة وجودة، مع وجود احتمال الفشل في بنائه بالدرجة الأولى.

**إدارة البيانات:**

إدارة البيانات هي إنشاء وتنفيذ البنى والسياسات والإجراءات التي تدير احتياجات دورة حياة البيانات الكاملة للمؤسسة.

## تحديات حوكمة البيانات Data Governance Challenges

تُعد التحديات المرتبطة بالبيانات السحابية والبيانات الضخمة من الأمور الشائعة التي تواجهها المؤسسات بخصوص حوكمة البيانات، فالخدمات السحابية وأنظمة البيانات الضخمة تستدعي متطلبات حوكمة جديدة. لقد كان تركيز برامج حوكمة البيانات حتى وقت قريب على البيانات المنظمة المخزنة في مركز البيانات، أما الآن فأصبح من الضروري التعامل مع الكثير من البيانات المنظمة وغير المنظمة وشبه المنظمة التي قد تظهر معًا في بيئات البيانات الضخمة، بالإضافة إلى تهديدات الخصوصية المرتبطة بأنظمة البيانات السحابية.

## من المسؤول عن حوكمة البيانات؟ Who is Responsible?

تضم عملية حوكمة البيانات مجموعة متنوعة من الأشخاص في معظم المؤسسات، ويتم تضمين المستخدمين النهائيين المطلعين على البيانات ذات العلاقة في أنظمة المؤسسة، وكذلك مدراء الأعمال، والمتخصصين في إدارة البيانات، وموظفي تقنية المعلومات، ويتحمل المسؤولية الرئيسة عن الحوكمة عادة رئيس قسم المعلومات (CIO) أو كبير مسؤولي البيانات (CDO) ومدير إدارة البيانات (DGM).

**مالك البيانات:**

شخص أو أشخاص مسؤولون عن بيانات محددة.

يُعدُّ رئيس قسم المعلومات أحد كبار المسؤولين التنفيذيين عن برنامج حوكمة البيانات، وتشمل مسؤولياته الحصول على الموافقة، والتمويل، والتوظيف في البرنامج، وكذلك تقديم المبادرات، وتقييم تطور البرنامج، والترويج له بفاعلية.

**مشرف بيانات:**

تتضمن مهام مشرف البيانات إدارة البيانات من خلال تنفيذ سياسات حوكمة البيانات والمحافظة عليها داخل المؤسسة.

فاعتمادًا على حجم المؤسسة، يتم تعيين مدير عام لإدارة البيانات ولقيادة وتنسيق مبادرة الحوكمة، حيث يتولى عقد الاجتماعات، وتنفيذ الدورات التدريبية، وتتبع مؤشرات الأداء الرئيسة، وإدارة الاتصالات الداخلية للمبادرة. ويعمل مدير إدارة البيانات مع مالكي البيانات والمسؤولين الذين يضمنون تطبيق ضوابط وقواعد حوكمة البيانات واتباع المستخدمين النهائيين لها.



# تمرينات

1

| حدد الجملة الصحيحة والجملة الخاطئة فيما يلي: | صحيحة | خاطئة |
|---|---|---|
| 1. تشير البيانات الضخمة إلى البيانات الكبيرة جدًا أو المعقدة التي لا يمكن معالجتها باستخدام الطرق التقليدية. | ○ | ○ |
| 2. من التقنيات الخمس التي تمكن من إدارة البيانات الضخمة السرعة والمصداقية ومستودع البيانات. | ○ | ○ |
| 3. يعتبر اكتشاف المعرفة عملية بسيطة لا تتطلب أي خطوات محددة. | ○ | ○ |
| 4. التخزين السحابي هو طريقة التخزين الوحيدة المستخدمة لكم البيانات الكبير مثل البيانات الضخمة. | ○ | ○ |
| 5. تُعدُّ قابلية التوسع والتكلفة المنخفضة لتحليل البيانات من المزايا العديدة لتخزين البيانات الضخمة سحابيًا. | ○ | ○ |
| 6. مستودع البيانات هو مستودع يتوافر سحابيًا عادةً لتخزين كميات هائلة من البيانات الأولية وغير المعالجة. | ○ | ○ |
| 7. الحوسبة في الذاكرة هي طريقة لتسهيل تحليل البيانات الضخمة لاعتمادها بصورة أساسية على ذاكرة الحاسب الرئيسة لتخزين البيانات. | ○ | ○ |
| 8. تشير بحيرة البيانات إلى قاعدة البيانات التي تخزن البيانات الحالية والتاريخية الناتجة عن أنظمة المعاملات الأساسية في المؤسسة. | ○ | ○ |
| 9. اختيار البيانات هو تحديد جزء من مجموعة البيانات الذي نريد استخدامه لعملية اكتشاف المعرفة. | ○ | ○ |
| 10. تمثيل المعرفة هو عملية استخراج البيانات من التحليلات من خلال الأنماط. | ○ | ○ |



2 اذكر ثلاثة أمثلة على استخدام البيانات الضخمة للمساعدة في الأعمال التجارية.

______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________

3 ابحث في الإنترنت عن مزودي خدمات الحوسبة السحابية الأكثر شهرة اليوم في السوق العالمية والمُستخدمة في تخزين ومعالجة البيانات الضخمة.

______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________

4 اشرح بإيجاز مدى مساعدة التخزين السحابي لنا في التعامل مع مشكلة تخزين الكم الهائل من البيانات الناتج عن البيانات الضخمة.

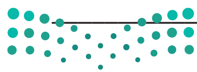
______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________



**5** البيانات الضخمة هي تطور حديث في تاريخ الحوسبة، فهل يمكنك تحديد عاملين ساهما في هذا النمو المفاجئ في جمع البيانات؟

______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________

**6** قارن بين التقنيات الثلاث لتخزين البيانات الضخمة، وإذا قمت بتطوير تطبيق يتطلب وصولًا سريعًا جدًا إلى البيانات، فما هي التقنية التي ستختارها؟

______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________

**7** لماذا يُعدُّ تقييم أنماط البيانات مهمًا في عملية التنقيب عن البيانات؟

______________________________________________________________

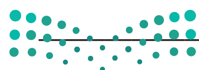
______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________

______________________________________________________________



**8** اشرح كيفية عمل قابلية التوسع في تخزين البيانات السحابية، ثم ابحث عن خدمتين لتخزين البيانات سحابيًا على شبكة الإنترنت.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**9** ما الغرض من حوكمة البيانات؟ وهل تعتبر حوكمة البيانات مرادفة لإدارة البيانات؟

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**10** ابحث في الإنترنت عن معلومات حول ضوابط أو قوانين إدارة البيانات الصحية في المملكة العربية السعودية، ثم اذكر العواقب المترتبة على تسرب بيانات أحد مرافق الرعاية الصحية؟

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________



**11** أنشئ تقريرًا عن تغير المناخ من خلال مقارنة بيانات الطقس التاريخية لدولتين من اختيارك، ثم حدّد أين ستبحث عن المعلومات على الإنترنت، واشرح العوامل المؤثرة وراء قرارك هذا.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**12** ما مخاوف الخصوصية التي يمكنك التفكير بها عند تعامل مؤسسة كبيرة مع البيانات الضخمة؟

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**13** هل لديك فكرة عن المعلومات التي تمتلكها شبكة التواصل الاجتماعي التي انضممت إليها حول عائلتك وأصدقائك؟ إذا كانت الإجابة نعم، اكتب قائمة قصيرة بهذه المعلومات.

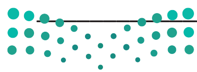
______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________




رابط الدرس الرقمي
www.ien.edu.sa

الدرس الثالث

# أساسيات علم البيانات

## علوم الرياضيات التي تحتاجها لتصبح عالم بيانات
## Mathematics Needed to Become a Data Scientist

تتطلب خوارزميات علم البيانات بالإضافة إلى تنفيذ التحليلات واكتشاف الأفكار من البيانات الموجودة معرفة رياضية، ففي حين أن الرياضيات لا تعتبر الأداة الوحيدة المطلوبة لعالم البيانات ولكنها من أهم تلك الأدوات. أحد العناصر الأكثر أهمية في سير عمل مشروع علم البيانات تحديد وفهم تحديات الأعمال وتحويلها إلى تحديات رياضية.

### الجبر الخطي Linear Algebra

يهتم الجبر الخطي بالمصفوفات والمتجهات، مما يُعد أمرًا مهمًا للغاية لأنه في نماذج علم البيانات والخوارزميات يتم تحويل جميع الأرقام والمعلومات إلى مصفوفات. ويتم استخدام تقنية أخرى ضرورية في معالجة البيانات الضخمة و ترتكز على الجبر الخطي، وهي تقنية تقليص الأبعاد. وكذلك تُعد **رؤية الحاسب** (Computer Vision) **والبرمجة اللغوية العصبية** (NLP) من مجالات علم البيانات التي تعتمد بشكل كبير على الجبر الخطي.

### الرياضيات المتقطعة Discrete Mathematics

تتخصص الرياضيات المتقطعة في طرق المنطق والاستنتاج، وهي جوانب أساسية في تصميم الخوارزميات وتُعدُّ أساس علم البيانات. ومن المجالات المهمة جدًا الخاصة بالرياضيات المتقطعة هي نظرية المخططات. تستخدم المخططات في نمذجة شبكات معقدة للغاية مثل شبكات تنظيم الجينات، وتعتبر دراسة هذه المخططات في علم البيانات مهمة جدًا للتقدم في بعض المجالات مثل الطب الدقيق وبيولوجيا الأنظمة وغيرها الكثير.

### الاحتمالات والإحصاء Probability and Statistics

عند إنشاء البيانات بعد تحليلها، يحتاج عالم البيانات إلى معرفة عملية بالإحصاء والاحتمالات لكي يتمكن من فهم وتفسير تلك البيانات. يستخدم علماء البيانات مقاييس مثل التباين والارتباط والانحراف المعياري على نطاق واسع للحصول على نظرة ثاقبة على العلاقات الأساسية لخصائص مجموعة البيانات.

### التفاضل والتكامل Calculus

يُعدُّ تمثيل النتائج من تحليل البيانات أمرًا بالغ الأهمية لتوفير معلومات مستنيرة من خلال إنشاء الرسوم والمخططات البيانية.

يعتبر التفاضل والتكامل جزءًا لا يتجزأ من الخوارزميات المستخدمة في العمليات الحسابية المعقدة المطلوبة لهذه العملية، ويتم استخدام خصائص مثل الاشتقاق الجزئي، والانحدار الخطي، والنزول الاشتقاقي على نطاق واسع في التطوير والتحسين وحساب الخسارة.



## البايثون في علم البيانات Python for Data Science

يفضل محترفو علم البيانات عامةً استخدام لغة البايثون في مشروعاتهم المختصة بهذا الجانب، فهي لغة برمجة عالية المستوى كائنية التوجه وسهلة التعلم، ومن السهل البدء في العمل على مشروع ما بحيث يمكنك بدء كتابة كود برمجي بسيط أو تصميم وتنفيذ حل باستخدام مبادئ البرمجة كائنية التوجه (OOP).

يوفر استخدام واجهات برمجة التطبيقات (APIs) والمكتبات القياسية الوصول إلى دوال قوية سهلة الاستخدام. توجد العديد من المكتبات الجاهزة للاستخدام في البايثون من قِبل المتخصصين في المؤسسات المختلفة تغطي مجموعة متنوعة من الاحتياجات مثل: استخراج البيانات، وإعداد البيانات وتحليلها، ومعالجة البيانات، والنمذجة التنبؤية، وتمثيل البيانات، وإعداد التقارير.

كذلك فإن مكتبات البايثون تدعم تطبيقات تعلّم الآلة ومتطلبات الذكاء الاصطناعي المتقدمة بما يتجاوز تطبيقات علم البيانات التقليدية.

**البايثون (Python):**
لغة برمجة عالية المستوى تستخدم لأغراض متعددة وقد اكتسبت شعبية متزايدة في علم البيانات وتعلّم الآلة.

## مقدمة إلى مفكرة جوبيتر
## Intro to Jupyter

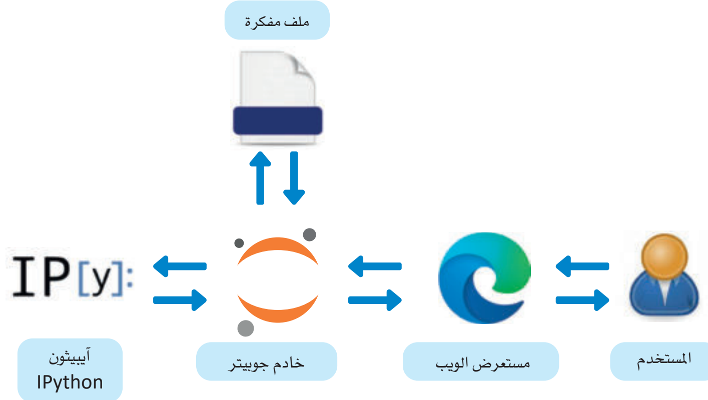
يمكن كتابة أوامر البايثون النصية في أحد بيئات التطوير المتكاملة (IDE) مثل **فيجول ستوديو كود** (Visual StudioCode) أو JetBrains PyCharm أو يمكنك كتابتها في **مفكرة جوبيتر** (Jupyter Notebook).

تعتبر مفكرة جوبيتر أحد تطبيقات الويب مفتوحة المصدر المستخدمة لتطوير وتقديم مشروعات علم البيانات باستخدام البايثون. تتيح البيئة التفاعلية لعلماء البيانات إنشاء مفكرات محوسبة، وتدمج مفكرة جوبيتر أوامر البايثون وتخرجها في مستند واحد يجمع بين التمثيلات والنص السردي والمعادلات الرياضية وأشكال البيانات الأخرى. بعد تثبيت البرنامج، يمكن تشغيله في متصفح الويب إما عبر الإنترنت أو على حاسب شخصي.

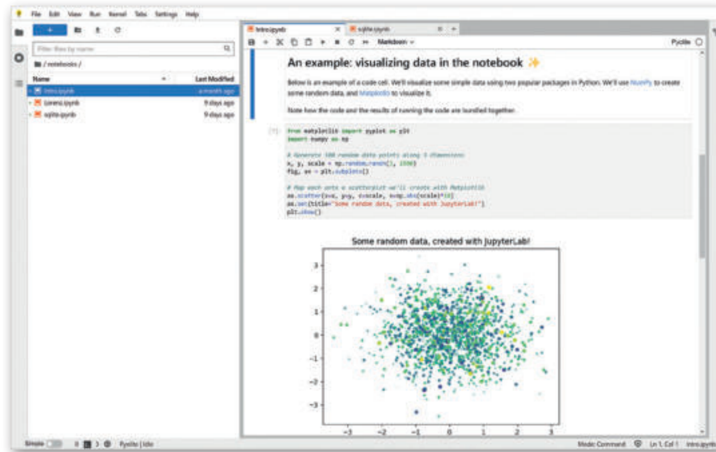
تدعم مفكرة جوبيتر إلى جانب البايثون أكثر من 100 لغة برمجة يطلق عليها اسم **أنوية** (kernels) في بيئة نظام مفكرة جوبيتر بما فيها Java و R وJulia وMATLAB وOctave وScheme وProcessing وScala وغيرها.

يقوم تطبيق مفكرة جوبيتر بتشغيل نواة IPython فقط ولكن يمكن تثبيت أنوية إضافية.

سنستخدم هذا البرنامج لتحليل البيانات الاستكشافية لاحقًا في هذا الكتاب، ويعتبر أحدث تطبيق قائم على الويب لاستخدام مفكرة جوبيتر هو تطبيق JupyterLab بحيث تعمل جميع المستندات بنفس الطريقة في كلتا البيئتين.



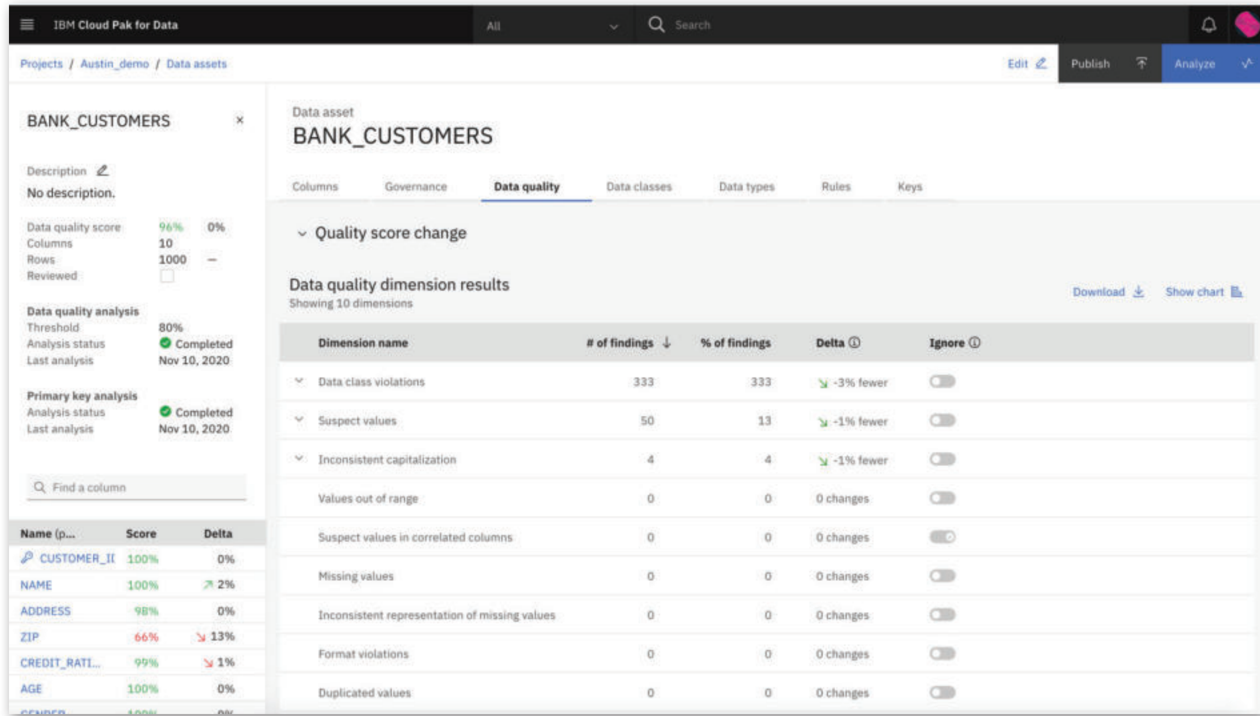
شكل 1.9: هيكلية مفكرة جوبيتر



شكل 1.10: نموذج لقطة شاشة من مفكرة جوبيتر

## أدوات علم البيانات Tools for Data Science

علم البيانات عملية معقدة تتطلب الكثير من الخطوات لتوفير حلول لعلم البيانات، ولكل خطوة من خطوات هذه العملية العديد من الأدوات لإنجاز المهمة المطلوبة. يعرض جدول 1.9 الأدوات الأكثر شيوعًا لكل خطوة في علم البيانات.




شكل 1.11: لقطة شاشة لنموذج بيانات من خلال IBM Cloud Pak.

**جدول 1.9: الأدوات الشائعة لعلم البيانات**

| | الغرض | أدوات البرمجيات |
|---|---|---|
| **تخزين البيانات** | قواعد البيانات حيث يتم تخزين البيانات. | MySQL و SQL Server و MongoDB و Neo4j. |
| **تحويل صيغة البيانات** | الأدوات التي تستعلم عن البيانات التي تريد تحليلها. | Python و SQL و Apache TinkerPop. |
| **النمذجة** | تحويل البيانات المطلوبة إلى نماذج مناسبة للتحليل. | Pandas و NumPy و Apache Spark. |
| **التحليل** | العملية التي تولّد الرؤى المطلوبة. | Tensorflow و PyTorch و IBM Watson و AWS و Sagemaker. |
| **التصوير** | تصوير النتائج في الشكل الأمثل. | Matplotlib و D3.js و R. |

## مهن علم البيانات Data Science jobs

علم البيانات هو أحد أسرع التخصصات المتعلقة بعلوم الحاسب نموًا والأكثر طلبًا في الوقت الراهن، وقد نشرت مؤسسة مسك مؤخرًا تقريرًا عن سوق العمل السعودي ووظائف المستقبل، حيث يركزّ التقرير على الوظائف الأكثر طلبًا حاليًا، وتبدو الفرص الوظيفية في علم البيانات واعدة بشكل خاص لا سيما الوظائف التي تدعم أهداف رؤية السعودية 2030.

**الجدول 1.10: المهن المتعلقة بعلم البيانات**

| | |
|---|---|
| **عالم بيانات** | مهمته هي إيجاد البيانات ومعالجتها وتحليلها للشركات والمؤسسات. يأخذ البيانات الأولية وغير المعالجة ثم يتم استخراج رؤى وأنماط من البيانات تساعد الشركات والمؤسسات على تحليل أدائها واتخاذ قرارات مهمة. |
| **مهندس تعلم آلة** | مسؤول عن تنفيذ حلول وأنظمة تعلّم الآلة لتطبيقات معينة. يجب أن يكون على دراية بهندسة البرمجيات وبالإحصاء ليتمكن من اختبار الحلول والحكم على صحة نماذج تعلّم الآلة الناتجة. |
| **متخصص تعلم آلة** | بينما يختص مهندسو تعلّم الآلة في تطبيق نماذج تعلّم الآلة، فإن متخصص تعلّم الآلة يُركز على إيجاد الخوارزميات الرياضية المحددة التي تنتج النماذج التي يمكن للمهندسين استخدامها بعد ذلك. |
| **معماري تطبيقات** | يقوم بتصميم نظم المعلومات للمؤسسات والشركات. |
| **معماري أنظمة المؤسسات الكبيرة** | يجمع بين المعرفة التجارية والتقنية، وعلى تواصل مستمر بين أصحاب الأعمال والإدارات الفنية، ويكلف بترجمة احتياجات بيانات الأعمال والمؤسسات إلى مواصفات وحلول تقنية تُرسل إلى الفرق الفنية. |
| **معماري بيانات** | مسؤول عن تخزين وسير المعلومات في شركة أو مؤسسة. يعمل مع علماء ومهندسي البيانات لتطوير طرق نقل البيانات بصورة مناسبة لإدخال مجموعة البيانات وتحليلها وإخراج النتائج. |
| **مهندس بيانات** | يساعد مهندس البيانات في بناء الإطار الرقمي لجمع البيانات وتخزينها ومعالجتها، والتي سيستخدمها علماء البيانات والمحللون في عملهم. |
| **معماري البنية التحتية** | دوره يختص في إدارة البنية التحتية حيث يتم تخزين البيانات ومعالجتها، ويأخذ بالاعتبار عوامل مثل خصوصية البيانات والحماية وأداء البنية التحتية على الخوادم حيث يتم تحليل البيانات، ولقد أصبحت مشروعات علم البيانات أكثر تعقيدًا بشكل عام، لذلك يحتاج مهندسو البنية التحتية إلى التأكد من اكتمال معالجة البيانات ضمن الجداول الزمنية المناسبة. |
| **محلل بيانات** | هو محترف يأخذ الرؤى من مجموعات البيانات المعالجة ويُنشئ التقارير والتصورات والتحليلات الأخرى المتنوعة والتي تتماشى مع الأهداف الرئيسة التي ينبغي للحل المبني على علم البيانات تحقيقها. |



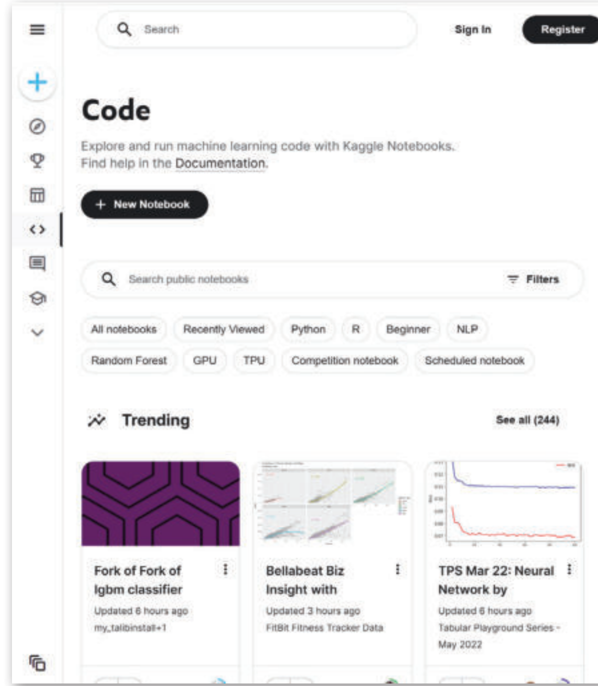
## مجتمعات علم البيانات عبر الإنترنت Data Science Online Communities

يرغب علماء البيانات في البقاء على اتصال مع أقرانهم في هذا المجال أو في المهن المماثلة لتعلم أفكار وطرق جديدة لأن منهجيات وتقنيات علم البيانات دائمة التغير. وتوفر الموارد عبر الإنترنت لعلماء البيانات الفرصة فقط في الحفاظ على وتيرة معينة، وهنا برزت الحاجة إلى وجود مجتمع من خبراء علم البيانات لدعم هذا العمل وظهور مجموعة متنوعة من المنتديات والمجموعات عبر الإنترنت والتي تمكنهم من الاتصال معًا وتطوير هذا المجال بكفاءة من خلال المشاركة في مجتمعاته عبر الإنترنت. تشاهد هنا أبرز المجتمعات رغم أنه يُعدُّ مجالًا مفتوحًا لظهور مجتمعات جديدة أخرى قد تكون ناجحة أيضًا.

### كاقل Kaggle

كاقل شركة تابعة لقوقل، وهي أكبر مجتمع لعلم البيانات يضم ملايين الأعضاء النشطين ومجموعة واسعة من الموارد، ويمكن لعلماء البيانات العثور على مجموعات البيانات العامة والمواد التعليمية وبيئة العمل المستندة إلى مجموعة النظراء لدعم عمل تحليل البيانات الخاص بهم.

https://www.kaggle.com




شكل 1.12: الصفحة الرئيسية لموقع Kaggle.com

### مجتمع بيانات IBM Data Community IBM

مجتمع بيانات IBM عبارة عن منتدى عبر الإنترنت به مدونات مخصصة لعلم البيانات. يستضيف الأوراق البحثية والبث عبر الإنترنت والعروض التقديمية التي يتم تحديثها مع تطور هذا المجال.

https://community.ibm.com/community/user/datascience

توجد المزيد من المجتمعات على الإنترنت المدعوم بعضها من قِبل الحكومات، ويدير البعض الآخر متطوعون.

يركز البعض بشكل أكبر على الجانب الاجتماعي من خلال الاجتماعات وجهًا لوجه، بينما يركز البعض الآخر على الأكواد البرمجية المطلوبة لمشروعات علم البيانات.

**جدول 1.11: المجتمعات عبر الإنترنت**

| | |
|---|---|
| مركز علم البيانات (Data science center) | https://www.datasciencecentral.com/ |
| تبادل المكدس (Stack Exchange) | https://datascience.stackexchange.com/ |
| مجتمع علم البيانات (Data science community) | https://dssberkeley.com/ |
| بيانات مدفوعة (Driven Data) | https://www.drivendata.org/ |
| مجتمع البيانات (Data Community DC) | https://www.datacommunitydc.org/ |
| مجتمع ردیت (Reddit Society) | https://www.reddit.com/r/datascience/ |

**تذكَّر دائمًا أن تتحقق من دقة البيانات أو الأكواد أو الأدوات عبر الإنترنت قبل استخدامها، تحقق أيضًا من تصاريح الاستخدام القانونية لكل مجموعة بيانات وحاول تنزيل أدوات البرامج مباشرة من ملفات مطوريها.**

# تمرينات

**1**

| حدد الجملة الصحيحة والجملة الخاطئة فيما يلي: | صحيحة | خاطئة |
|---|---|---|
| 1. يتم تحويل جميع الأرقام والمعلومات إلى مصفوفات في نماذج وخوارزميات تعلّم الآلة. | ○ | ○ |
| 2. لكي يتمكن عالم البيانات من فهم وتفسير البيانات عند إنشاءها، فإنه يحتاج إلى معرفة عملية بالإحصاء والاحتمالات. | ○ | ○ |
| 3. تتخصص الرياضيات المتقطعة في طرق المنطق والاستنتاج، وهي جوانب أساسية في تصميم الخوارزمية والتي تُعدُّ الأساس لتعلّم الآلة. | ○ | ○ |
| 4. بعض المجتمعات على الإنترنت مدعومة من قبل الحكومات، والبعض الآخر يديره متطوعون. | ○ | ○ |
| 5. مهندس الأنظمة هو الشخص الذي يصمم أنظمة المعلومات للمؤسسات والشركات. | ○ | ○ |
| 6. عالم البيانات هو محترف يأخذ الرؤى من مجموعات البيانات التي تمت معالجتها ويقوم بإنشاء تقارير وتصورات وتحليلات أخرى متنوعة تتماشى مع الأهداف الرئيسة التي تحتاج إلى حل مبني على علم البيانات لتحقيقها. | ○ | ○ |
| 7. محلل البيانات هو محترف مسؤول عن تخزين وتدفق المعلومات في شركة أو مؤسسة، ويعمل مع علماء ومهندسي البيانات لبناء طرق نقل البيانات المناسبة لإدخال مجموعة البيانات وتحليلها وإخراج النتائج. | ○ | ○ |



**2** وضّح كيف تُمكّن لغة البايثون من مساعدة المتخصص في علم البيانات.

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

**3** وضِّح كيف يمكن لجوبيتر نوت بوك مساعدة المتخصص في علم البيانات.

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

**4** عدِّد أهم أدوات علم البيانات، وكيف يمكن لكل أداة المساهمة في كل خطوة من خطوات علم البيانات؟

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________



5 لماذا يعتبر فهم الإحصاء مهارة أساسية لعالم البيانات؟ وهل يمكنك التفكير في مثال يتضمن تحليل البيانات؟

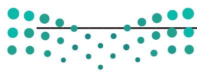
______________________________________________
______________________________________________
______________________________________________
______________________________________________
______________________________________________

6 البايثون هي لغة برمجة متعددة الاستخدامات، فهل تُعدُّ كافية لمشروعات علوم البيانات؟

______________________________________________
______________________________________________
______________________________________________
______________________________________________
______________________________________________
______________________________________________

7 ابحث عن ثلاث مكتبات البايثون تحظى بشعبية كبيرة بين علماء البيانات على الإنترنت، ثم اشرح باختصار سبب ذلك.

______________________________________________
______________________________________________
______________________________________________
______________________________________________
______________________________________________



8 قارن بين بيئة التطوير المتكاملة و بيئة جوبيتر نوت بوك، ثم اذكر الاختلافات الرئيسة بينهما؟

___

___

___

9 على فرض أنك عالم بيانات جديد تتقن لغة البايثون، فما الأدوات الأخرى التي تحتاجها من أدوات علم البيانات؟

___

___

___

10 يوجد في هذا الدرس قائمة بالمهن المتعلقة بعلم البيانات، فأي منها تفضّل؟ ولماذا؟ وما التحديات التي تعتقد بأنك ستواجهها في هذه المهنة؟

___

___

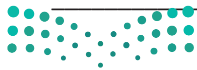
___

11 قم بزيارة أحد مجتمعات علم البيانات عبر الإنترنت وابحث عن دورة تدريبية بسيطة للدراسة الذاتية لتعزيز معرفتك بعلم البيانات، ثم قم بتقييم مدى ملاءمة الدورة التدريبية لمستوى معرفتك.

___

___

___



# المشروع

1

توفّر الشبكات الاجتماعية كميات هائلة من المعلومات كل يوم. حدِّد ثلاثة إجراءات يومية تنتج بيانات خاصة مفيدة بهذه الشبكات.

2

فكّر في الأمور التالية على وجه التحديد:

ما أنواع البيانات التي يتم جمعها؟

هل جميع هذه البيانات متاحة للجمهور؟

3

قم بإعداد شرائح لعرض مخاطر انتهاك الخصوصية المتعلقة بوسائل التواصل الاجتماعي، ووضِّح كيف يمكن حماية المستخدمين، واذكر أفضل الخطوات التي يمكن اتخاذها لتتجنب أن تصبح بياناتك معلومات مفيدة يمكن للآخرين استغلالها؟



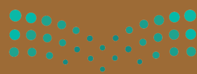


# ماذا تعلمت

- المقصود بعلم البيانات.
- الفرق بين البيانات والمعلومات والمعرفة.
- كيفية التمييز بين علم البيانات وذكاء الأعمال والذكاء الاصطناعي.
- تعريف دورة حياة علوم البيانات.
- المقصود بالبيانات الضخمة.
- كيفية مساهمة البايثون أو الأدوات الأخرى في علم البيانات.

## المصطلحات الرئيسة

| | | | |
|---|---|---|---|
| Artificial Intelligence | ذكاء اصطناعي | Data Visualization | تصوير البيانات |
| Big Data | البيانات الضخمة | Data Warehouse | مستودع البيانات |
| Business Intelligence | ذكاء الأعمال | Exploratory Data Analysis | التحليل الاستكشافي للبيانات |
| Cloud Storage | التخزين السحابي | Information | معلومات |
| Data | البيانات | In-Memory Computing | الحوسبة في الذاكرة |
| Data Analysis | تحليل البيانات | Jupyter Notebook | مفكرة جوبيتر |
| Data Analyst | محلل البيانات | Knowledge | المعرفة |
| Data Cleaning | تنظيف البيانات | Python | البايثون |
| Data Collection | جمع البيانات | Raw Data | بيانات أولية |
| Data Lake | بحيرة بيانات | Value | القيمة |
| Data Mining | التنقيب في البيانات | Variety | التنوع |
| Data Preparation | تجهيز البيانات | Velocity | السرعة |
| Data Science | علم البيانات | Veracity | الموثوقية |
| Data Science Life Cycle | دورة حياة علم البيانات | Volume | الحجم |
| Data Scientist | عالم البيانات | | |



وزارة التعليم
Ministry of Education
2022 - 1444

# 2. جمع البيانات والتحقق من صحتها



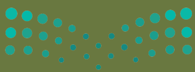
سيتعرَّف الطالب في هذه الوحدة على مفاهيم جمع البيانات وطرق التحقق من صحتها.

ويشمل ذلك معنى جمع البيانات وأنماطها المختلفة و مصادرها. وسيتم التطرق إلى موضوع ترميز البيانات والتركيز على ميزات وعيوب هذه الطرق، وفي الختام سيتعرف الطالب على خطوات التحقق من صحة البيانات وأنماطها.

## أهداف التعلم

بنهاية هذه الوحدة سيكون الطالب قادرًا على أن:

- يعرِّف مصطلح جمع البيانات.
- يصنِّف مصادر البيانات.
- يصنِّف معايير جودة المعلومات.
- يعرِّف مفهوم منصات البيانات المفتوحة.
- يتعرَّف على أهمية الحصول على التصاريح القانونية المناسبة قبل البدء بجمع البيانات.
- يحدد الأنماط المختلفة للبيانات.
- يعرِّف المقصود بترميز البيانات.
- يُطبِّق خطوات التحقق من صحة البيانات.
- يصنِّف الطرق المتنوعة للتحقق من صحة إدخال البيانات.




رابط الدرس الرقمي
www.ien.edu.sa

الدرس الأول

# جمع البيانات

## جمع البيانات Data Collection

تُعدُّ مرحلة **جمع البيانات** (Data Collection) أهم مراحل البحث العلمي أو دراسة ظاهرة معينة، وتُمثل هذه المرحلة عملية جمع الحقائق والأرقام والمصطلحات للمتغيرات المستهدفة في ذلك البحث أو الدراسة. يمكن جمع البيانات باستخدام أجهزة مختلفة مثل أجهزة الاستشعار وتسجيل البيانات.

تتطلب عملية جمع البيانات فهمًا عميقًا للمُعامِلات قيد الدراسة، بالإضافة إلى التخطيط والعمل الدؤوب للحصول على بيانات عالية الجودة، حيث تسهم البيانات عالية الجودة في إجراء التحليل المناسب لأداء المهام على الوجه المطلوب وذلك لاستخراج معلومات مفيدة حول الظاهرة قيد الدراسة.

تختلف طرق جمع البيانات باختلاف نوع البيانات، ولكن يظل التحقق من صحة ودقة الخطوات المتعلقة بجمع البيانات أمرًا هامًا.

**جمع البيانات (Data Collection):**
هي عملية جمع القراءات أو الحقائق وقياسها، وتشمل الحصول على البيانات وتسميتها وتحسينها.


الشكل 2.1: مهندس يجمع بيانات حول حالة الطقس

### مثال

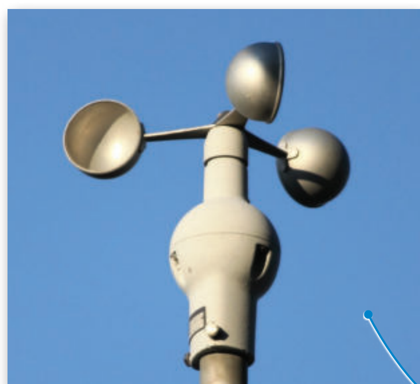
**تعد معرفة حالة الطقس من أهم الأمور المتعلقة بالاستعداد للسفر. يمكن استخدام العديد من الأجهزة لدراسة العوامل المتعلقة بالمناخ، ومنها أجهزة قياس درجة الحرارة ومقياس شدة وسرعة الرياح ومقياس درجة الرطوبة.**
**البيانات التي يتم جمعها من هذه الأجهزة هي قيم درجة الحرارة، وحساب سرعة الرياح واتجاهاتها، ونسبة بخار الماء في الجو.**

## مصادر البيانات Sources of Data

يوجد تصنيفان أساسيان لمصادر البيانات: **مصادر البيانات الرئيسة** (Primary data sources)، **ومصادر البيانات الثانوية** (Secondary data sources).

### مصادر البيانات الرئيسة Primary Data Source

يحتوي مصدر البيانات الرئيسة على بيانات لم تجمع من قبل ويمكن جمعها من خلال أجهزة الإستشعار وأجهزة تسجيل البيانات وكذلك من خلال الاستبانات. من الأمثلة التي تدل على مصادر البيانات الرئيسة: مقياس درجة الحرارة الذي يجمع بيانات درجة حرارة الهواء، ومقياس سرعة الرياح الذي يقيس سرعة الرياح، وكذلك فإن إجراء استبانة للعملاء حول طبيعة الطقس الذي يفضلونه عند السفر للخارج يُعدُّ مصدرًا من مصادر البيانات الرئيسة.



الشكل 2.2: مقياس سرعة الرياح

**مقياس سرعة الرياح هو جهاز يستخدم لحساب السرعة في حركة الرياح، حيث تدفع الرياح الناتجة عن حركة الهواء الزعانف الثلاثة للمقياس مما يتسبب في دورانها، مما يؤدي إلى دوران المحور المركزي الذي يتصل بمولد كهربائي. تفعّل الطاقة الناتجة عن هذا المولد، عداد كهربائي يتم معايرته لعرض سرعة الرياح.**

### مصادر البيانات الثانوية Secondary Data Source

تحصل على هذا النوع من البيانات من خلال استخدام بيانات أخرى من مصدر البيانات الرئيس. فعلى سبيل المثال، يمكنك استخدام بيانات درجة حرارة الهواء وسرعة الرياح من جهازي استشعار مختلفين للحصول على بيانات لمُعامل آخر يسمى **درجة حرارة الرياح الباردة** (wind-chill temperature). يمكن حساب تلك الدرجة باستخدام صيغة حساب درجة حرارة الرياح الباردة، والتي تتمثل في ضرب سرعة الرياح بالمعامل 0.7 ثم طرح الناتج من درجة حرارة الهواء التي تم تسجيلها. بعبارة أخرى يمكنك أولًا استخدام مقاييس درجة الحرارة ومقاييس الرياح كمصادر بيانات أولية لحساب بيانات درجة الحرارة وسرعة الرياح، ثم يمكنك استخدام صيغة حساب درجة البرودة كمصدر بيانات ثانوي من أجل الحصول على بيانات درجة البرودة.




الشكل 2.3: الموقع الإلكتروني الخاص بتوقعات حالة الطقس

**الجدول 2.1: أوجه الاختلاف بين مصادر البيانات الرئيسة والثانوية**

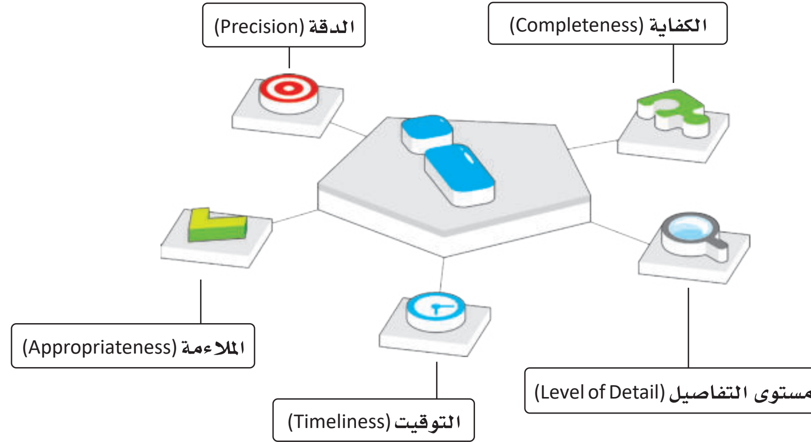
| | مصادر البيانات الرئيسة | مصادر البيانات الثانوية |
|---|---|---|
| **الأصالة** | يتم جمعها مباشرة من مصادرها الأصلية. | لا تعتبر بيانات أصلية لأنها تعتمد على بيانات أخرى تم جمعها من قبل آخرين. |
| **الشكل** | تكون بصورة خام وغير معرّفة. | تكون بصورة منظمة ومعالجة. |
| **الدقة** | أكثر دقة لكونها بيانات حالية. | أقل دقة لأنها مرتبطة بأحداث ماضية. |
| **المصدر** | يتم جمعها بواسطة أجهزة الاستشعار، الاستبانات، المقابلات والتجارب وغيرها. | يتم جمعها من الكتب والمجلات والمستندات والصفحات الإلكترونية والمدونات وغير ذلك. |
| **التكلفة** | عالية التكلفة وتتطلب الكثير من الوقت لجمعها. | أقل تكلفة وأقل استهلاكًا للوقت. |

## مصادر البيانات الداخلية والخارجية Internal and External Data Sources

يمكن تصنيف مصادر البيانات إلى مصادر داخلية وخارجية. تعكس مصادر البيانات الداخلية تلك البيانات التي يتم جمعها داخليًا من قبل جهة العمل أو الجهة المسؤولة عن جمع البيانات، بينما البيانات الخارجية هي التي يتم تحصيلها من خارج مجال جهة العمل. وعلى سبيل المثال، تعتبر البيانات التي يتم جمعها من جهاز استشعار تابع لجامعة أو مؤسسة علمية بيانات داخلية، بينما تعتبر البيانات التي يتم الحصول عليها من مؤسسات أخرى أو أفراد أو من مصادر خارج الجامعة بيانات خارجية.

## جودة المعلومات Information Quality

عندما تتم معالجة البيانات أو تنظيمها أو تقديمها في سياق معين فإنها تسمى معلومات، وتوصف قيمتها لاستخدام معين بأنها "جودة المعلومات". تُعد جودة المعلومات عاملًا مهمًا في قيمة تلك المعلومات ومدى مصداقيتها للاستخدام في اتخاذ القرارات. ومع زيادة جمع وحفظ البيانات، أصبحت جودة المعلومات الناتجة عن معالجتها ذات أهمية كبيرة ومتزايدة. يساعد ضمان جودة المعلومات على التحديد الدقيق للاحتياجات الفعلية لتنفيذ المشروعات، وكذلك توجيه الخدمات بشكل فعال، وزيادة الكفاءة في العمل. وبالمقارنة يمكن أن تتسبب المعلومات غير الدقيقة في حدوث خلل في الأعمال، وتقلل من الكفاءة وتؤدي إلى التأخير في إنجاز المشروعات. يمكن التحقق من جودة المعلومات من خلال معايير محددة تسمى **معايير الجودة** (Quality standards)، وهي موضحة في الشكل التالي:



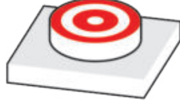

الشكل 2.4: معايير جودة المعلومات

| فيما يلي بعض الأسئلة التي يمكن أن تساعدك في التحقق من دقة المعلومات: |
| --- |
| هل يمكن التحقق من الحقائق والإحصاءات والمعلومات من مصادر أخرى؟ |
| هل يمكن تكرار التجربة التي تم الحصول على البيانات منها؟ وهل لها نفس النتائج؟ |
| ما مصدر المعلومات؟ |
| لماذا تم إنشاء المعلومات؟ |
| هل تبدو المعلومات دقيقة بناءً على معرفتك وخبرتك؟ |
| هل تتضمن المعلومات أخطاء إملائية أو أحرفًا في غير مكانها؟ وهل الاقتباسات التي تم الاستشهاد بها صحيحة؟ |

قبل جمع أي نوع من المعلومات من خلال موقع إلكتروني وقبل الشروع في الخطوة التالية في علم البيانات (وهي خطوة المعرفة)، يجب عليك التحقق من جودة المعلومات التي نحن على وشك الحصول عليها من الموقع. إذا كانت المعلومات ليست موثوقة، فلا يمكنك بكل تأكيد المضي قدمًا في استخراج المعرفة، وهذا يعني أنه يجب التحقق من المعلومات باتباع معايير الجودة الخمسة التالية:

تكمن أهمية معايير الجودة الخمس هذه في أنها تساعدك على التحقق من موثوقية كل المعلومات التي تجدها في المواقع الإلكترونية.

### أولًا: الدقة Accuracy



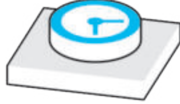
الدقة هي التأكد من صحة المعلومات، ويجب أن تكون المعلومات صحيحة لكي تُعدّ عالية الجودة. ويتم التحقق من دقتها من خلال المصادر الموثوقة، حيث يسهل التأكد من خلوها من الأخطاء الحسابية أو اللغوية أو غيرها من الأخطاء.

### ثانيًا: الملاءمة Appropriateness



يجب أن تكون المعلومات مرتبطة بموضوعك أو بالسؤال البحثي، فكلما كانت المعلومات متعلقة بما تبحث عنه، كلما كانت ملاءمتها أفضل. يؤدي الحصول على معلومات إضافية لا تتعلق بما تبحث عنه إلى إضاعة الوقت في محاولة العثور على المعلومات المطلوبة.

### ثالثًا: التوقيت Timeliness



يُعدُّ تاريخ نشر المعلومات جزءًا مهمًا، حيث يوضح مدى حداثة المعلومات ومناسبتها لموضوع البحث، ولذلك يجب التأكد من الحصول على آخر تحديث للمعلومات، ويجب عليك التحقق دائمًا من مصدر المعلومات.

### رابعًا: مستوى التفاصيل Level of detail



تُحدد جودة المعلومات أيضا من خلال النظر إلى مستوى التفاصيل التي تقدمها تلك المعلومات.

يُعدُّ وجود الكثير من التفاصيل على المعلومات المطلوبة أمرًا صعبًا، بينما تؤدي قلة التفاصيل إلى صعوبة فهم المعلومات. ويعتمد مستوى التفاصيل على المشكلة ودراستها، فإذا كانت بسيطة فإنك تحتاج إلى مستوى قليل من التفاصيل، وإذا كانت معقدة فإنك تحتاج إلى مستوى عالٍ من التفاصيل.

إن المقدار الصحيح من المعلومات هو معيار أساسي لجودتها.

### خامسًا: الكفاية Completeness

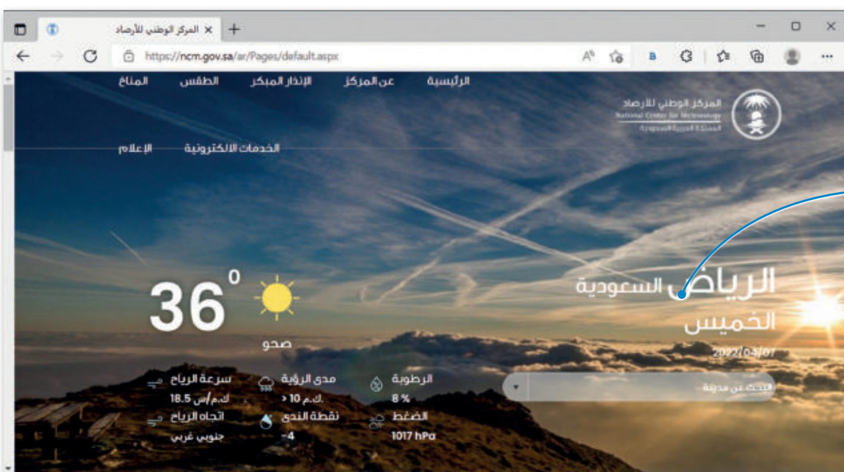


تُعدُّ كفاية المعلومات مقياسًا مهمًا للشمولية المطلوبة للتأكد من أن المعلومات المقدمة تعطي صورة كاملة عن الواقع. إن عدم الحصول على جميع المعلومات المطلوبة يعني أنك لن تتمكن من استخدامها بشكل صحيح، مما يعني أن جودة تلك المعلومات ضعيفة وغير كاملة ولا يمكن اتخاذ القرارات الصحيحة بناءً على تلك المعلومات.



**عند استخدام معيار التوقيت يجب مراعاة الأمور التالية :**

| |
|---|
| التحقق من تواريخ المصادر المستخدمة. |
| التحقق من التواريخ الخاصة بالحقوق الفكرية مثل تسجيل العلامات التجارية المسجلة، وحقوق النسخ والنشر، وبراءات الاختراع، والأسرار التجارية. |
| التحقق من تاريخ مراجعة أو تحرير المعلومات. |
| التحقق من تاريخ النشر. |




يُعدُّ الموقع الإلكتروني الخاص بالمركز الوطني للأرصاد مثالًا للعثور على المعلومات التي تلبي المعايير الخمسة المحددة لجودة المعلومات الموضحة أعلاه.

الشكل 2.5: مثال على مصدر معلومات

## منصات البيانات المفتوحة Open Data Platforms

منصات البيانات المفتوحة هي منصات تدعم وصول المستخدمين لمجموعات البيانات المفتوحة. تقدم منصات البيانات المفتوحة النموذجية بيانات المؤسسة التي تستضيف هذه المنصة.

تستضيف الجهات الحكومية أو المؤسسات غير الربحية منصات البيانات المفتوحة التي تتيح وصول الجمهور إلى البيانات. وعلى وجه التحديد فإنها تقوم بجمع البيانات وتنظيمها باستمرار من مجموعة متنوعة من القطاعات العامة. ويمكن استخدام مجموعات البيانات هذه دون تكلفة مالية أو تحديات تقنية للوصول إليها. كما يمكن إعادة استخدام البيانات المفتوحة وإعادة توزيعها مع مراعاة المتطلبات التي يفرضها ترخيص البيانات. ويمكن أيضًا استخدامها من قبل مواطني الدول الأخرى. قد تقدم الشركات أيضًا بيانات مفتوحة من خلال برامج المسؤولية الاجتماعية للشركات. هذه بعض الاستخدامات الشائعة لمنصات البيانات المفتوحة:

- الشفافية في إعداد الموازنة الحكومية والإنفاق على خدمات الدولة.
- إحصائيات أداء المؤسسات الحكومية.
- بيانات أداء مختلف القطاعات العامة للدولة كالتعليم والرعاية الصحية والنقل، والتي يمكن استخدامها في إجراء الأبحاث التي تُبرز الأداء العام داخل هذه الدولة.
- يمكن دمج مجموعات البيانات معًا واستخدامها في تطبيقات أخرى.

يمكن العثور على منصة البيانات المفتوحة الحكومية في المملكة العربية السعودية على عنوان الموقع الإلكتروني:

https://data.gov.sa/




الشكل 2.6: البوابة الوطنية للبيانات المفتوحة في المملكة العربية السعودية



## خصوصية البيانات Data Privacy

مصطلح "البيانات الشخصية" يُطلق على البيانات المتعلقة بالأشخاص وهي تلك البيانات التي من خلالها يتم تحديد هوية شخص معيّن، مثل اسم الشخص وعائلته ورقم هاتفه ورقم الهوية. أدّت عملية توسع وسائل التواصل الإلكتروني إلى زيادة في المخاطر المتعلقة بالخصوصية والتي عليك حماية نفسك منها، وتتضمن خصوصية البيانات قدرة الشخص على أن يقرر بنفسه متى وكيف وإلى أي مدى يمكن مشاركة معلوماته الشخصية مع الآخرين أو تزويد الآخرين بتلك المعلومات.

## الأذونات القانونية لجمع واستخدام البيانات
## Legal Permissions to Collect and Use Data

يتطلب جمع البيانات واستخدامها في المشروعات البحثية توفر التصاريح القانونية المناسبة، وعادة ما يقوم **مجلس المراجعة المؤسسية** (IRB) في المؤسسات بمراجعة مقترحات المشروعات البحثية قبل البدء بها لتحديد ما إذا كان المشروع يتبع المبادئ الأخلاقية واللوائح القانونية. تختلف هذه التصاريح في طبيعتها، ولكن العاملين الرئيسين اللذين يجب أخذهما بعين الاعتبار هما الموقع الذي سيتم فيه تخزين البيانات، وموقع المستفيدين الذين سيتعاملون مع تلك البيانات. تحتاج الشركات والمؤسسات إلى التأكد من أن الخدمات التي تجمع البيانات وتستخدمها متوافقة قانونيًا مع قوانين بلدانها.

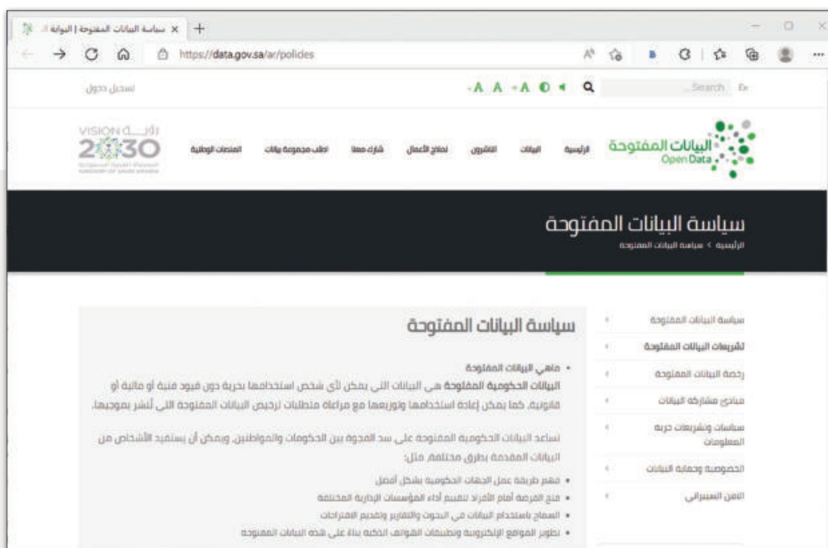
**مثال**

**يجب على مستخدمي البيانات التي يتم استضافتها على منصة البيانات المفتوحة في المملكة العربية السعودية استخدام هذه البيانات وفق شروط سياسة البيانات المفتوحة** https://data.gov.sa/ar/policies

## البحث الموجه ومقارنة البيانات Targeted Research and Data Comparison

يتم استخدام البحث الموجه للتركيز على قضايا معينة ظهرت من البحث الرئيس. فعلى سبيل المثال إذا استخدمت قيم درجة الحرارة والرياح للتنبؤ بالطقس في مدينة ما، ثم لاحظت أن مناطق معينة من هذه المدينة سجلت قيمًا قصوى لدرجات الحرارة، يتعين عليك في هذه الحالة إجراء بحث حول تلك المناطق لتقييم تأثير العوامل الأخرى على درجة الحرارة فيها.

يتم إجراء مقارنة البيانات عند وجود أكثر من مجموعة بيانات مسجلة لنفس المنطقة في فترات زمنية متماثلة. على سبيل المثال، قد يكون لديك مجموعة بيانات لقيم درجات الحرارة المسجلة لمدينة جدة في مارس 2021 ومجموعة أخرى مسجلة في مارس 2022. وبوجود مجموعتي البيانات هاتين، يمكنك بسهولة إجراء مقارنة البيانات من أجل اكتشاف التغيرات في درجات الحرارة أو التغيرات خلال سنوات.




الشكل 2.7: سياسة البيانات المفتوحة

# تمرينات

1

| حدد الجملة الصحيحة والجملة الخاطئة فيما يلي: | صحيحة | خاطئة |
|---|---|---|
| 1. جمع البيانات هو عملية جمع وقياس القراءات أو الحقائق. | ⚪ | ⚪ |
| 2. هناك نوعان من التصنيفات الأساسية لمصادر جمع البيانات: الرئيسة والثانوية. | ⚪ | ⚪ |
| 3. يعتبر تاريخ نشر المعلومات من أهم معايير جودة المعلومات. | ⚪ | ⚪ |
| 4. الملاءمة تعني أنه كلما كانت المعلومات ليست ذات صلة بما يتم البحث عنه، كانت جودتها أسوأ. | ⚪ | ⚪ |
| 5. تعتبر مستويات التفاصيل والدقة معايير جودة للمعلومات. | ⚪ | ⚪ |
| 6. تساعدك معايير الجودة الخمس في التحقق من موثوقية المعلومات. | ⚪ | ⚪ |
| 7. ليس للحكومة سلطة على منصات البيانات المفتوحة. | ⚪ | ⚪ |
| 8. تختلف التصاريح القانونية لجمع واستخدام البيانات بناءًا على عدة متغيرات. | ⚪ | ⚪ |
| 9. يتم استخدام البحث الموجه عندما نريد التركيز على قضايا محددة ظهرت من بحثنا الأساسي. | ⚪ | ⚪ |
| 10. يمكن إجراء مقارنة البيانات عندما يكون لديك أكثر من مجموعة بيانات واحدة مع بيانات مسجلة من نفس المنطقة ومن فترات زمنية مماثلة. | ⚪ | ⚪ |



**2** قارن بصورة مختصرة بين مصادر البيانات الرئيسة والثانوية.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**3** صِفْ باختصار كل سمة جودة يمكن استخدامها للتحقق من جودة المعلومات.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**4** أعطِ مثالًا على البحث الموجه ومقارنة البيانات.

______________________________________________

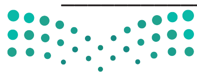
______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________



5 قارن بين مصادر البيانات الرئيسة والثانوية في حالة الطقس مع طرح بعض الأمثلة.

______________________________________________________________________
______________________________________________________________________
______________________________________________________________________
______________________________________________________________________
______________________________________________________________________
______________________________________________________________________

6 زُر منصة البيانات المفتوحة **data.gov.sa** وابحث عن المعلومات المتعلقة بالحصول على تصاريح باستخدام تلك البيانات. هل هناك أي استثناء؟

______________________________________________________________________
______________________________________________________________________
______________________________________________________________________
______________________________________________________________________
______________________________________________________________________
______________________________________________________________________



وزارة التعليم
Ministry of Education
2022 - 1444

**7** ابحث على الإنترنت عن منصات البيانات المفتوحة في البلدان الأخرى، وهل يمكنك العثور على معلومات شخصية عبر هذه المنصات؟

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

**8** حدد موقعين على الإنترنت أحدهما حكومي والآخر خاص، ثم قارن بين جودة المعلومات بينهما بناءً على المعايير الخمس التي تعرفت عليها.

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________




رابط الدرس الرقمي
www.ien.edu.sa

الدرس الثاني

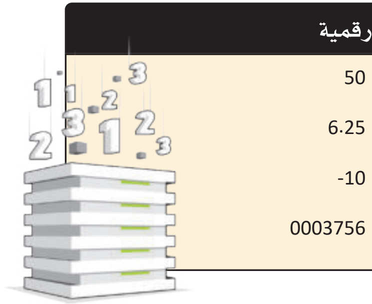
# أنواع البيانات

## البيانات الرقمية والبيانات الفئوية Numerical Data and Categorical Data

بعد أن تعرفت على مفهوم البيانات؛ الآن ستتعرف على أنماطها المختلفة. يمكن أن تتخذ البيانات عدة أشكال: مثل عدد زوار حدث معين، أو مدة تلك الزيارة، أو غير ذلك. في الدراسات البحثية، ثمة نوعان أساسيان من البيانات وهما: البيانات الرقمية والبيانات الفئوية.

## البيانات الرقمية Numerical Data

تتكون البيانات الرقمية من حقائق قابلة للقياس، كعدد الفعاليات المختلفة في مدينة ما. ويمكن للبيانات الرقمية أن تكون منفصلة أو تسلسلية.



الشكل 2.8: البيانات الرقمية

### البيانات المنفصلة Discrete Data

تمثل البيانات المنفصلة عناصر قابلة للعد يمكنها أن تأخذ قيمًا معينة فقط، مثل عدد الطلبة في الفصل.

### البيانات التسلسلية Continuous Data

البيانات التسلسلية تمثل بيانات يمكن قياسها ويمكنها أن تأخذ أي قيمة، مثل طول الشخص.

## البيانات الفئوية Categorical Data

البيانات الفئوية هي نوع بيانات غير كمية، يمكن تخزينها وتحديدها بناءً على الأسماء أو التسميات المعطاة لها. ويمكن أن توصف هذه البيانات على أنها اسمية أو ترتيبية.



الشكل 2.9: البيانات الفئوية

### البيانات الاسمية Nominal Data

يتم تعريف البيانات الاسمية على أنها بيانات تُستخدم لتسمية المتغيرات أو توسيمها، دون تحديد قيمة كمية لها أو وضعها في ترتيب معين. على سبيل المثال: يمكن تصنيف نتائج اختبارٍ ما اسميًا: "ناجح" أو "راسب".

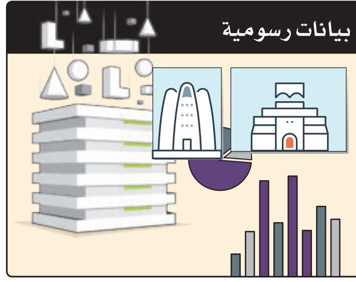
### البيانات الترتيبية Ordinal Data

البيانات الترتيبية هي نوع من البيانات الفئوية التي تتبع ترتيبًا معينًا. يتم فرز مجموعات البيانات الترتيبية وفقًا لنظام تصنيف محدد. على سبيل المثال: يمكن فرز نتائج اختبار ما بترتيب تنازلي حسب الدرجة: A ثم B ثم C ثم D ثم F.



## البيانات الرسومية أو بيانات مقاطع الصوت أو الفيديو
## Graphical, Video and Audio Data Types

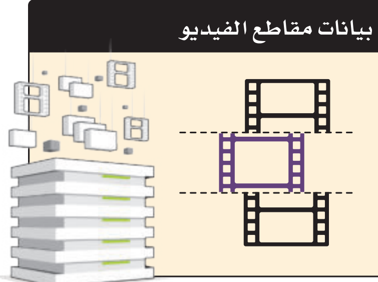
على الرغم من أن البيانات عادة ما تكون في شكل أبجدي رقمي (نصوص وأرقام ورموز)، إلا أنها قد تتكون من صور أو مقاطع صوتية أو مقاطع فيديو. فيما يلي بعض الأنواع الأخرى من البيانات:



الشكل 2.10: البيانات الرسومية

### البيانات الرسومية Graphical Data

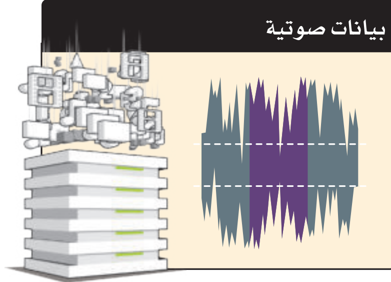
تتكون البيانات الرسومية من مخططات ورسوم بيانية وغير ذلك. على سبيل المثال: مجموعة الصور الخاصة بالمعالم السياحية لمنطقة محددة، أو الرسم البياني الخاص بأعداد الزوار لأحد الأماكن السياحية في المملكة العربية السعودية.



الشكل 2.11: بيانات الصور ومقاطع الفيديو

### بيانات الصور ومقاطع الفيديو Image and Video Data

قد تكون الصورة الرقمية صورة أو رسمًا توضيحيًا يتم تمثيله بوحدات البكسل أو بالمتجهات. تتكون بيانات الفيديو من سلسلة من الصور المتحركة والصوت ومن الأمثلة عليها الإعلان التلفزيوني لحملة سياحية، أو مقطع فيديو عن بوليفارد الرياض، أو بث فيديو مباشر من قناة القرآن الكريم أو مراسم الحج في المملكة العربية السعودية وما إلى ذلك.



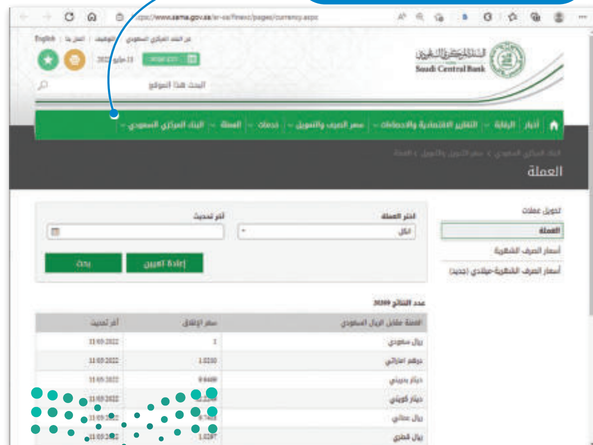
الشكل 2.12: البيانات الصوتية

### البيانات الصوتية Audio Data

تتكون البيانات الصوتية من الأصوات وكذلك المؤثرات الصوتية المختلفة كالتسجيلات الصوتية الإرشادية في المتاحف والأماكن السياحية المختلفة في المملكة العربية السعودية.

## البيانات الثابتة والمتغيرة Static and Dynamic Data

قد تبقى البيانات على حالتها الأصلية بعد تسجيلها وقد تتغير أحيانًا، ولذلك يمكن تمثيل البيانات بشكل ثابت أو متغير.



الشكل 2.13: موقع البنك المركزي السعودي

### البيانات الثابتة Static Data

البيانات الثابتة هي البيانات التي لا تتغير بعد تسجيلها، مثلًا البيانات التي تحتويها مجلة مطبوعة حول الأماكن السياحية في المملكة العربية السعودية، فبمجرد طباعة المجلة لا يمكن تغيير المعلومات فيها.

### البيانات المتغيرة (الديناميكية) Dynamic Data

البيانات المتغيرة هي البيانات التي قد تتغير بعد تسجيلها ويجب تحديثها باستمرار. على سبيل المثال بيانات موقع إلكتروني يحتوي على معلومات عن الأماكن السياحية في المملكة العربية السعودية، حيث يمكن تحديثها عند الحاجة.

## ترميز البيانات Data Coding

إن البيانات التي تم الحصول عليها من التجارب أو من خلال الدراسات الاستقصائية هي بيانات في صورتها الأولية وفي الغالب تحتاج إلى ترميز. تتيح عملية الترميز تنظيم البيانات وترتيبها بطريقة محددة وذلك باستخدام رموز مختلفة مثل الأرقام أو الحروف أو الكلمات القصيرة، والتي تصف سياقًا معينًا أو تعبّر عن عبارة أو فقرة بأكملها، وفيما يلي أمثلة من الحياة اليومية حيث تُستَخدَم الرموز لتمثيل البيانات:

### رموز المطارات Airport Codes

وضع **اتحاد النقل الجوي الدولي** (IATA) رمزًا مكونًا من ثلاثة حروف يحدد العديد من المطارات حول العالم. يمكنك البحث عن تذاكر الطيران على الإنترنت باستخدام رمز المطار، ويمكنك ملاحظة الرمز كذلك على بطاقات الأمتعة التي يتم إلصاقها بالحقائب عند تسليمها في مكاتب تسجيل المغادرة بالمطار استعدادًا لركوب الطائرة؛ مما يتيح شحن الأمتعة للوجهة المقصودة، كما يظهر الرمز على بطاقات الأمتعة المرفقة في مكاتب تسجيل الوصول بالمطار؛ لتوفير السلامة في حالة فقدان الأمتعة.

**الجدول 2.2: رموز المطارات**

| الرمز | التفسير |
|---|---|
| DMM | مطار الملك فهد الدولي |
| JED | مطار الملك عبد العزيز الدولي |
| RUH | مطار الملك خالد الدولي |

### رموز العملات Currency Codes

لكل بلد في أنحاء العالم عملة خاصة به، وتستخدم رموز العملات بدلًا من اسم العملة كاختصارات متعارف عليها عند التعاملات المالية.

الشكل 2.14: رموز العملات

**الجدول 2.3: رموز العملات**

| الرمز | التفسير |
|---|---|
| SAR | الريال السعودي |
| USD | الدولار الأمريكي |
| EUR | اليورو |

**الجدول 2.4: مزايا ترميز البيانات**

| المزايا | التفسير |
|---|---|
| إدخال أسرع للبيانات | من الأسهل كتابة الرمز RUH بدلًا من كتابة مطار الملك خالد الدولي. |
| تأخذ مساحة أقل | من الصعب كتابة الاسم الكامل للبلد على اللوحات أو على وسائل النقل العام مثل سيارات الأجرة والحافلات، ولكن يصبح الأمر سهلًا مع رموز تسجيل المركبات الدولية. |
| تسريع عملية البحث عن البيانات | لكل منطقة رمزها الخاص. يُستخدم هذا الرمز للبحث عن عنوان حسب رمز المنطقة، ورقم الشارع، والمبنى، كما يستخدمه مكتب البريد لتسهيل توزيع رسائل البريد. |

**الجدول 2.5: عيوب ترميز البيانات**

| العيوب | التفسير |
| --- | --- |
| معنى غامض للبيانات | قد تجد صعوبة في التمييز بين الرموز المتشابهة. |
| صعوبة فهم الترميز | قد يكون من الصعب تفسير أو تذكر معنى الرمز. |
| الرموز المستخدمة قد تُستنفد | مثلا قد يكون عدد العناصر المراد ترميزها كبيرا جدا بحيث لا تكفي مجموعة الأحرف لترميزها، فتُدمج الأرقام والحروف أو تستخدم الأرقام الطويلة وهذا يعقّد عملية الترميز، مثل ترميز المنتجات الاستهلاكية في المتاجر. |

### الرموز الشريطية Barcodes

ترى الرموز الشريطية في كل مكان حولك وبشكل يومي، فمثلا تجدها مطبوعة على التذاكر الإلكترونية والمنتجات في محلات البقالة وغيرها. الرمز الشريطي هو ملصق به خطوط سوداء رفيعة إلى جانب مجموعة متنوعة من الأرقام. تُستخدم في تنظيم المعلومات وفهرستها أو وضع علامة على أسعار المنتجات.




الشكل 2.15: مثال على الرموز الشرطية

### رقم الكتاب المعياري الدولي ISBN (International Standard Book Number)

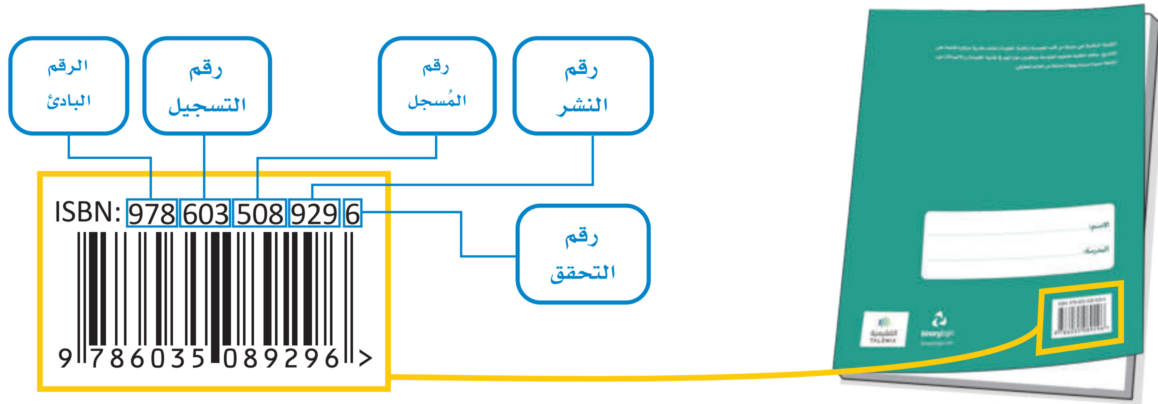
يوجد رقم على غلاف معظم الكتب (مثل الكتيب الإرشادي) أعلى الرمز الشريطي يُسمى **رقم الكتاب المعياري الدولي** (ISBN) وهو رقم فريد يستخدمه الناشرون والمكتبات ومحلات بيع الكتب لتحديد عناوين الكتب وإصداراتها. يتكون رقم الكتاب المعياري الدولي من ثلاثة عشر خانة عشرية ويقسم إلى خمس مجموعات متتالية من الأرقام.

**الجدول 2.6: هيكل ISBN المكون من 13 رقمًا**

| مجموعة الأرقام | الوصف |
|---|---|
| رقم البادئ | يتكون من ثلاثة أرقام إما تكون 978 أو 979. بالنسبة لكتب المملكة العربية السعودية فإن الرقم البادئ هو 978. |
| رقم التسجيل | يتكون من رقم واحد أو قد يصل إلى خمسة أرقام ويستخدم كمعرِّف للبلد. وبالنسبة لكتب المملكة العربية السعودية فإن رقم التسجيل هو 603 أو 9960. |
| رقم المُسجل | قد يصل إلى سبعة أرقام ويستخدم لتحديد الناشر. |
| رقم النشر | قد يصل إلى ستة أرقام ويستخدم لتحديد إصدار محدد، والشكل لعنوان محدد. |
| رقم التحقق | يتكون من رقم واحد ويوضع دائمًا في النهاية، ويستخدم للتحقق من صحة باقي الأرقام. |

## مثال

فيما يلي مثال على رمز **ISBN**، حيث يشير كل عنصر إلى جزء معين من المعلومات حول الدليل.




الشكل 2.16: مثال على ISBN المكون من 13 رقمًا

## رموز الاستجابة السريعة QR Codes

**رموز الاستجابة السريعة** (Quick Response QR) هو بمثابة الجيل الثاني من **الرمز الشريطي** (barcode)، والذي يتكون من خطوط سوداء متجاورة ومختلفة السُمك ويحتوي على مزيد من المعلومات. قد يشير رمز الاستجابة السريعة إلى محتوى إلكتروني مثل: المواقع الإلكترونية، أو مقاطع الفيديو، أو الملفات الرقمية، ويمكن قراءة هذا الرمز باستخدام كاميرات الهواتف الذكية.



يشير رمز الاستجابة السريعة هذا إلى رابط الموقع الإلكتروني:
https://www.visitsaudi.com

الشكل 2.17: مثال على رموز الاستجابة السريعة

# تمرينات

1

| حدد الجملة الصحيحة والجملة الخاطئة فيما يلي: | صحيحة | خاطئة |
|---|---|---|
| 1. البيانات الفئوية هي نوع كمي من البيانات. | ○ | ○ |
| 2. يتم تعريف البيانات الترتيبية على أنها البيانات المستخدمة لتسمية المتغيرات أو توسيمها. | ○ | ○ |
| 3. تمثل البيانات المنفصلة عناصر قابلة للعد ويمكن أن تأخذ قيمًا معينة فقط. | ○ | ○ |
| 4. رموز المطارات ورموز العملات هي أمثلة على ترميز البيانات. | ○ | ○ |
| 5. البيانات الثابتة هي البيانات التي قد تتغير بعد تسجيلها ويجب تحديثها باستمرار. | ○ | ○ |
| 6. البيانات المتغيرة (الديناميكية) هي البيانات التي لا تتغير بعد تسجيلها. | ○ | ○ |
| 7. غالبًا ما يتم إجراء تشفير البيانات على البيانات في شكلها الأولي، والتي يتم الحصول عليها من التجارب أو من خلال الاستطلاعات. | ○ | ○ |
| 8. يتكون رقم ISBN من 10 مجموعات متتالية من الأرقام. | ○ | ○ |
| 9. الرمز الشريطي عبارة عن ملصق به خطوط سوداء رفيعة، مع مجموعة متنوعة من الأرقام. | ○ | ○ |
| 10. يتكون رمز الاستجابة السريعة من مربعات سوداء تحتوي على معلومات. | ○ | ○ |



2 صِفْ باختصار البيانات الثابتة والبيانات المتغيرة.

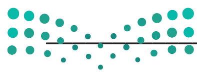
______________________________________________
______________________________________________
______________________________________________
______________________________________________
______________________________________________
______________________________________________

3 اذكر بعض الأمثلة على منتجات مميزة برموز شريطية أو رموز الاستجابة السريعة.

______________________________________________
______________________________________________
______________________________________________
______________________________________________
______________________________________________
______________________________________________

4 وضِّح باختصار المقصود بترميز البيانات.

______________________________________________
______________________________________________
______________________________________________
______________________________________________
______________________________________________
______________________________________________
______________________________________________



**5** ابحث عن موقع الكتروني يمكن من خلاله إنشاء رموز QR مجانية، وأنشئ رمز QR للصفحة الرئيسة لذلك الموقع، ولصفحة ويب أخرى لموقع الكتروني من اختيارك. هل يمكنك ملاحظة الاختلافات في المربعات السوداء لكل رمز QR؟

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**6** تحتفظ المنظمة الدولية للتوحيد القياسي بالقائمة الرسمية لرموز الدول من خلال معيار (ISO 3166). ابحث عن رموز دول مجلس التعاون الخليجي (GCC) المكونة من حرفين. هل يمكنك إعطاء أمثلة على استخدام هذه الرموز؟

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**7** ابحث عن رمز ISBN لهذا الكتاب، وهل يمكنك تحديد أرقام الدولة والناشر؟

______________________________________________

______________________________________________

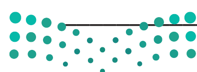
______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________



الدرس الثالث

# التحقق من صحة إدخال البيانات

رابط الدرس الرقمي



www.ien.edu.sa

## التحقق من صحة إدخال البيانات Data Entry Validation

يشير مفهوم التحقق من صحة إدخال البيانات إلى أي نشاط يتحقق من أن البيانات المدخلة تنبثق من مجموعة من القيم المعتمدة، وتتوافق مع القواعد المقبولة للبيانات، وقد تتبع تلك البيانات بعض العمليات والإجراءات التصحيحية، وتهدف عملية التحقق من صحة البيانات إلى ضمان الدقة والجودة، وتنفذ من خلال إنشاء عدة فحوصات لضمان الاتساق المنطقي للبيانات المدخلة والمخزنة. ويتم قبول البيانات في حال توافقها مع القوانين، ويتم تجاهل ورفض البيانات في حال عدم الالتزام بالقوانين.

**التحقق من صحة البيانات:**

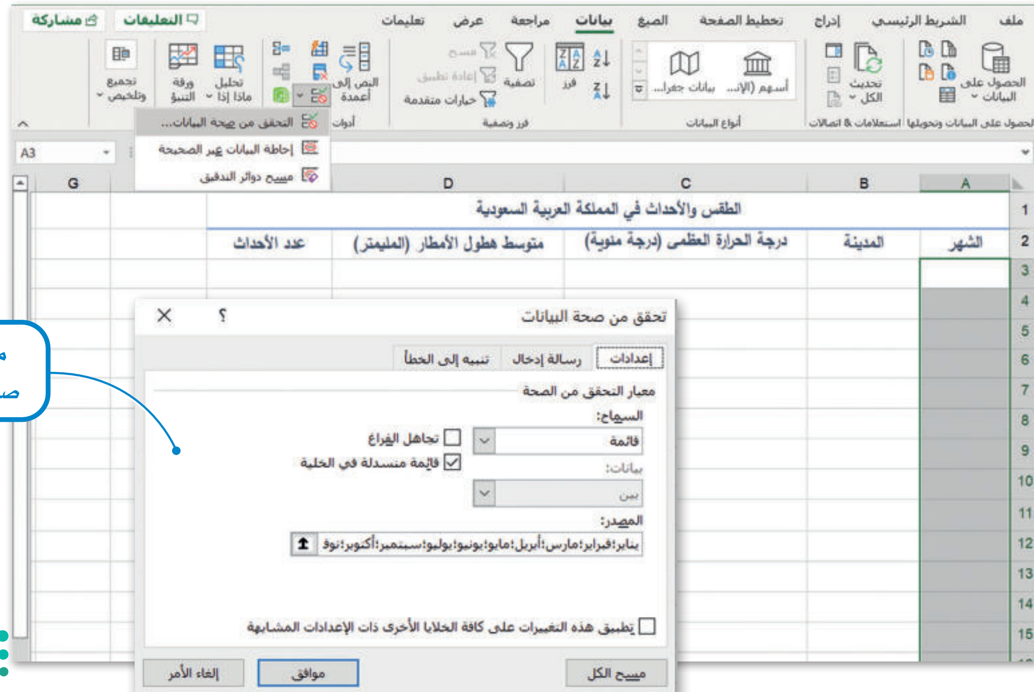
هي عملية ضمان دقة وجودة البيانات ويتم تنفيذها عن طريق عدة فحوصات للتأكد من التناسق المنطقي للبيانات المدخلة والمخزنة.

**على سبيل المثال، يمكن أن يتراوح النطاق المقبول لقيم درجة حرارة الهواء المسجلة من جهاز استشعار درجة الحرارة من سالب 89.2 درجة مئوية (أدنى درجة حرارة) إلى 58 درجة مئوية (أعلى درجة حرارة)، ويجب ألا يسجل جهاز الاستشعار قيم درجة حرارة الهواء مثل 100 درجة مئوية، حيث يشير ظهور مثل هذا النوع من البيانات في نتائج جهاز الاستشعار المسجلة إلى حدوث خلل ويجب رفض القيمة.**

### أنواع التحقق من صحة إدخال البيانات Types of Data Entry Validation

يوجد العديد من أنواع التحقق التي يمكن استخدامها للتأكد من صحة البيانات المدخلة، كما يمكن استخدام برامج مثل **مايكروسوفت إكسل** (Microsoft Excel) للتحقق من صحة البيانات المدخلة.

تتبع معظم أساليب التحقق من صحة إدخال البيانات طريقة أو أكثر من طرق التحقق أدناه للتأكد من أن البيانات صحيحة قبل تخزينها. وأكثر هذه الطرق شيوعًا هي: التحقق من البحث، والتحقق من التواجد، والتحقق من الطول، والتحقق من النطاق، والتحقق من الصيغة، والتحقق من النوع.




الشكل 2.18: مثال على التحقق من صحة البيانات في إكسل

**الجدول 2.7: أنواع التحقق من صحة البيانات:**

| النمط | الوصف |
|---|---|
| التحقق من البحث (Lookup check) | يساعد على تقليل الأخطاء باستخدام قائمة محدودة النطاق تحتوي على قيم مدخلة مسبقًا. فبدلًا من كتابة اسم المطار، يمكن لموظف شركة الطيران اختيار المطار من قائمة تحتوي على رموز المطارات حول العالم. هذا النوع من التحقق يساعد على ادخال البيانات بسرعة وبنسبة أخطاء أقل. |
| التحقق من التواجد (Presence check) | يجعل عملية الإدخال إلزامية في الخلية مما يضمن عدم تركها فارغة. يجب إدخال البيانات الهامة؛ وإلا فإن هناك مشكلة في تكامل البيانات، مثلًا لا يمكن ترك الحقول الخاصة باسم الشخص واسم عائلته فارغة. |
| التحقق من الطول (Length check) | يهدف إلى التأكد من أن الرموز والحروف تُدخل بنطاق طول محدد. مثلًا أرقام ISBN أو رموز العملات أو رموز البلدان ISO لها طول ثابت مكون من 13 و 3 و 2 رقمًا أو حرفًا، على التوالي. |
| التحقق من النطاق (Range check) | يُستخدم للتأكد من أن الأرقام المُدخلة تقع ضمن نطاق معين ويشمل حدين هما: الحد الأقصى (Maximum limit) والحد الأدنى (Minimum limit)، فمثلا إذا كنت تريد إدخال عمر الشخص، فيجب على النظام قبول الأرقام الموجبة بحد أعلى 140 فقط، وأي شيء آخر يتجاوز هذا النطاق سَيُعدُّ بيانات غير صالحة. |
| التحقق من الصيغة (Format check) | يُستخدم للتأكد من أن البيانات تأتي بصيغة محددة مسبقًا ولن يسمح بأي صيغة أخرى يتم إدخالها في الخلية. تكون عمليات التحقق هذه مفيدة عند استخدام حقول البيانات للرموز البريدية أو أرقام الهواتف. في كلتا الحالتين، يتوقع النظام تنسيق بيانات محدد للغاية. |
| التحقق من النوع (Type check) | يضمن إدخال المستخدمين لنوع القيمة الصحيح في حقل محدد. فمثلا إذا تم تمييز حقل البيانات على أنه حقل رقمي، فلن تتمكن من تخزين قيمة نصية. |

## رقم التحقق Check Digit

يتم استخدام رقم التحقق للتأكد من صحة البيانات في الأرقام ذات الطول الثابت. يمكن أن يتكون رقم التحقق من خانة أو خانتين إضافيتين للرقم، ويستخدم رقم التحقق على نطاق واسع في التطبيقات المصرفية حيث تحتاج الحسابات المصرفية إلى التحقق من أرقام التعريف التي يتم إدخالها يدويًا وتخضع لفحص بسيط للكشف عن الخطأ، وتحسب الخوارزمية رقم التحقق من الأرقام الأخرى للرقم الذي تم إدخاله وتقارنه بالأرقام المكتوبة.

إذا كان هناك رقم مفقود أو تمت كتابته بشكل خطأ، فسيعرض النظام تحذيرًا للتحقق من صحة البيانات. تتضمن أرقام ISBN وأرقام التذاكر ومجموعة كبيرة من الرموز الشريطية رقم تحقق. في السنوات الأخيرة تضاءلت أهمية رقم التحقق نظرًا لإدخال البيانات بشكل آلي عبر الماسحات والكاميرات والطرق الآلية الأخرى.




الشكل 2.19: رقم التحقق في ISBN



## مثال على التحقق من صحة البيانات Data Validation Example

يوجد العديد من التطبيقات والبرامج التي يمكن استخدامها للتحقق من صحة البيانات المدخلة، مثل برنامج مايكروسوفت إكسل. ستنشئ ورقة بيانات في برنامج مايكروسوفت إكسل من أجل تطبيق التحقق من صحة البيانات التي تعلمتها في الجزء النظري.

افترض أنك تعمل كوكيل سياحي، وتريد عمل حملة سياحية لمدينتي الرياض وجدة حيث ستقام فعاليات أو أحداث مهمة طوال العام، ومن العوامل المهمة في تنظيم الحملة السياحية أن تكون على دراية بأحوال الطقس في كل مدينة؛ من أجل اشعار السياح للاستعداد لهذه الظروف والاستمتاع بالحدث بشكلٍ أفضل.

بصفتك وكيل سياحي، زرت **موقع المركز الوطني للأرصاد** (https://ncm.gov.sa) وحمَّلت بيانات درجات الحرارة وهطول الأمطار لمدينتي جدة والرياض. افتح برنامج مايكروسوفت إكسل وأنشئ ورقة تسمى "الطقس والأحداث في المملكة العربية السعودية" لإدخال هذه البيانات. قم بإنشاء خمسة أعمدة بعنوان الأشهر والمدينة ودرجة الحرارة العظمى (درجة مئوية)، ومتوسط هطول الأمطار (المليمتر) وعدد الأحداث. عليك برمجة القيم المحددة للتحقق من صحة البيانات في كل عمود، وبعد ذلك سيتعين عليك إدخال البيانات التي حصلت عليها من موقع المركز الوطني للأرصاد كما في الجدول التالي:

**الجدول 2.8: الطقس والأحداث في المملكة العربية السعودية**

| الأشهر | المدينة | درجة الحرارة العظمى (درجة مئوية) | متوسط هطول الأمطار (المليمتر) | عدد الأحداث |
|---|---|---|---|---|
| يناير | جدة | 28.8 | 12.50 | 2 |
| يناير | الرياض | 20.7 | 14.80 | 5 |
| فبراير | جدة | 29.8 | 3.30 | 1 |
| فبراير | الرياض | 23.7 | 8.30 | 8 |
| مارس | جدة | 25.5 | 2.60 | 1 |
| مارس | الرياض | 28 | 19.90 | 7 |
| أبريل | الرياض | 33.6 | 23.70 | 1 |
| مايو | جدة | 30.7 | 0.10 | 1 |
| مايو | الرياض | 39.5 | 5.60 | 1 |
| يونيو | جدة | 38.2 | 0.00 | 1 |
| يوليو | جدة | 39.4 | 0.40 | 2 |
| سبتمبر | الرياض | 32.8 | 0.00 | 4 |
| أكتوبر | الرياض | 27.5 | 1.50 | 4 |
| نوفمبر | جدة | 27.6 | 27.10 | 1 |
| نوفمبر | جدة | 20.4 | 20.00 | 3 |

يمكن تلخيص إجراءات التحقق من صحة البيانات التي سيتم اتباعها كما في المخطط التالي:

| الإجراء | الوصف |
|---|---|
| التحقق من البحث | القيم المحددة مسبقا من عمود "الشهر". |
| التحقق من التواجد | التأكد من أن كل خلية لها قيمة مسجلة. |
| التحقق من الطول | يجب أن تتكون أسماء المدن المسجلة من ثلاثة إلى ستة حروف. |
| التحقق من النطاق | يجب أن تتراوح قيم درجة الحرارة المسجلة من 20 إلى 45 درجة مئوية. |
| التحقق من الصيغة | يجب أن تكون قيم هطول الأمطار مسجلة بصيغة رقم عشري. |
| التحقق من النوع | يجب أن تكون القيم رقمية فقط. |

الشكل 2.20: إجراءات التحقق من صحة البيانات

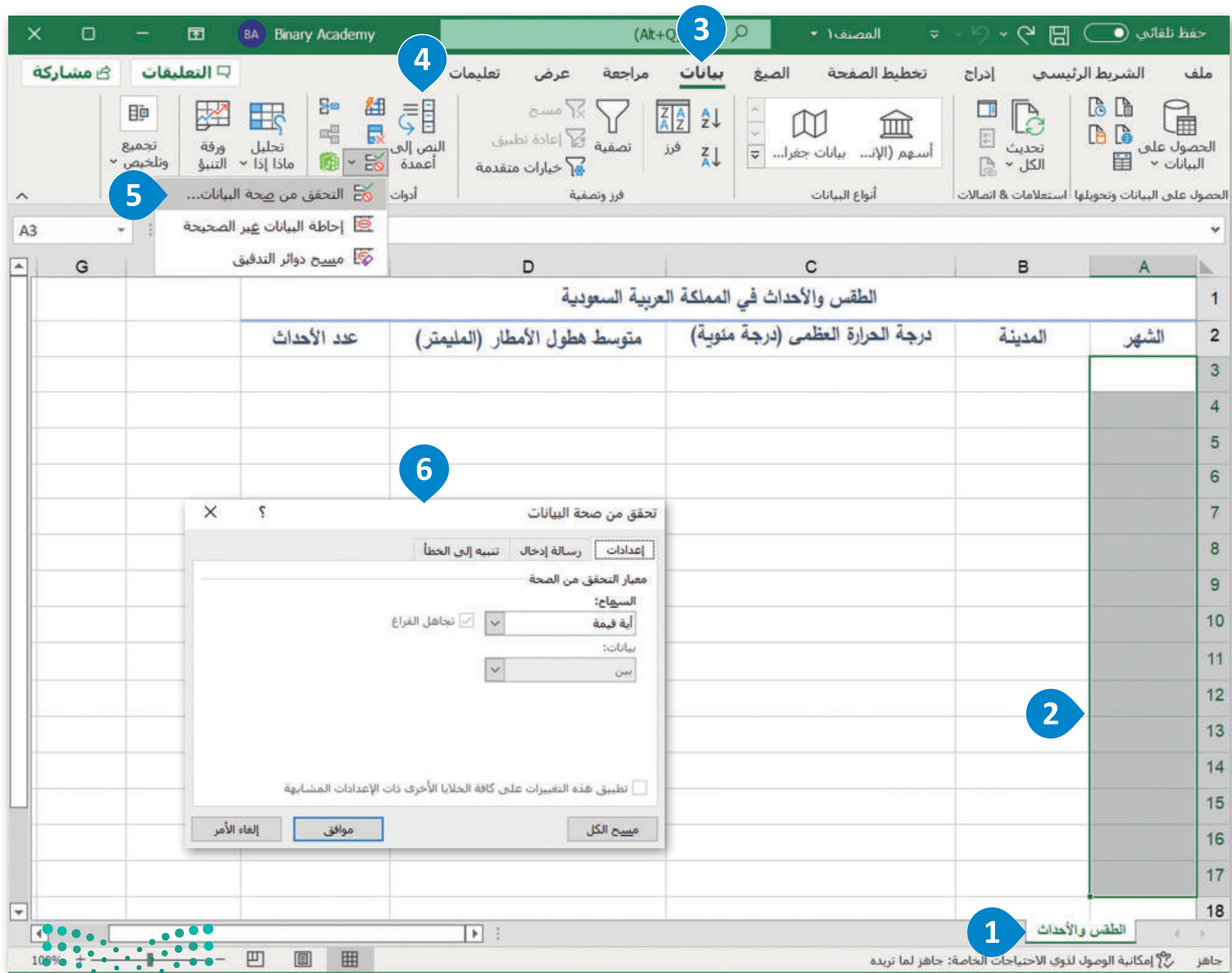


## التحقق من صحة البحث والتواجد في إكسل Lookup and Presence Validation Check in Excel

لتنفيذ التحقق من صحة البحث والتواجد على العمود الأول بعنوان "الشهر"، ستقوم بإضافة جميع أسماء الأشهر كقيم مُحددة ومرتبة مسبقًا لكي يستطيع المستخدم إدخالها بمجرد البحث عن كل شهر من القائمة. ستستخدم نافذة التحقق من صحة البيانات بحيث لا يستطيع البرنامج قبول الخلايا الفارغة والتي تعني أن الخلايا بدون قيمة.

**لبدء التحقق من صحة البيانات في إكسل:**

- انتقل إلى ورقة "**الطقس والأحداث في المملكة العربية السعودية**" في إكسل. 1
- حدد الخلايا من **A3** إلى **A17**. 2
- ثم اختر علامة التبويب **بيانات** (Data). 3
- من مجموعة **أدوات البيانات** (Data tools)، 4 اضغط على نافذة **التحقق من صحة البيانات** (Data Validation). 5
- ستظهر نافذة **التحقق من صحة البيانات** (Data Validation). 6




الشكل 2.21: بدء التحقق من صحة البيانات في إكسل

## لتطبيق التحقق من صحة البحث والتواجد في إكسل:

< من نافذة **تحقق من صحة البيانات** (Data Validation)، 1
اختر علامة التبويب **إعدادات** (Settings). 2
< من مربع **السماح** (Allow)، اختر **قائمة** (List). 3
< من مربع **المصدر** (Source)، اكتب أسماء الأشهر وافصل بين كل واحد والآخر بفاصلة منقوطة "؛". 4
< ألغ تحديد خيار **تجاهل الفراغ** (Ignore Blank). 5




يتم التحقق من البحث عندما تضيف أسماء الأشهر في مربع المصدر (Source).

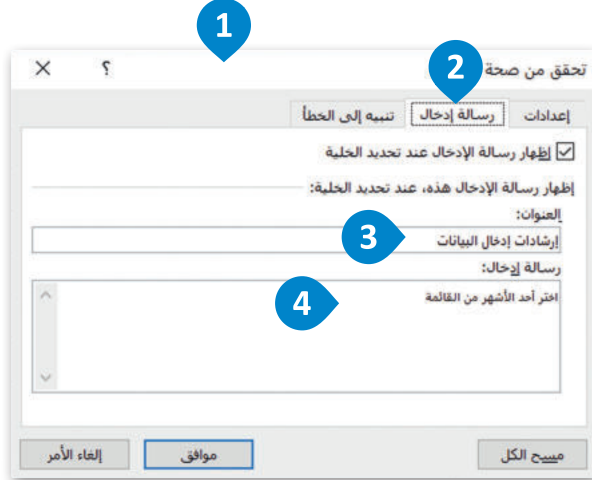
يتم التحقق من التواجد عندما تلغي تحديد الخيار تجاهل الفراغ (Ignore Blank).

الشكل 2.22: تطبيق التحقق من صحة التواجد والبحث في إكسل

**من المهم جدا منع المستخدم من إدخال نوع خاطئ من البيانات، ولهذا الغرض تضع رسائل تنبيه إلى الخطأ وإدخال غير صالح حتى يتم تذكير المستخدم أثناء عملية إدخال البيانات.**

## لتعيين رسالة إدخال غير صالح:

< من نافذة **تحقق من صحة البيانات** (Data Validation)، 1
اختر علامة التبويب **رسالة إدخال** (Input Message). 2
< في مربع **العنوان:** (Title:) اكتب "إرشادات إدخال البيانات". 3
< في مربع **رسالة إدخال** (Input message)، اكتب "اختر أحد الأشهر من القائمة". 4




الشكل 2.23: تعيين رسالة إدخال غير صالح

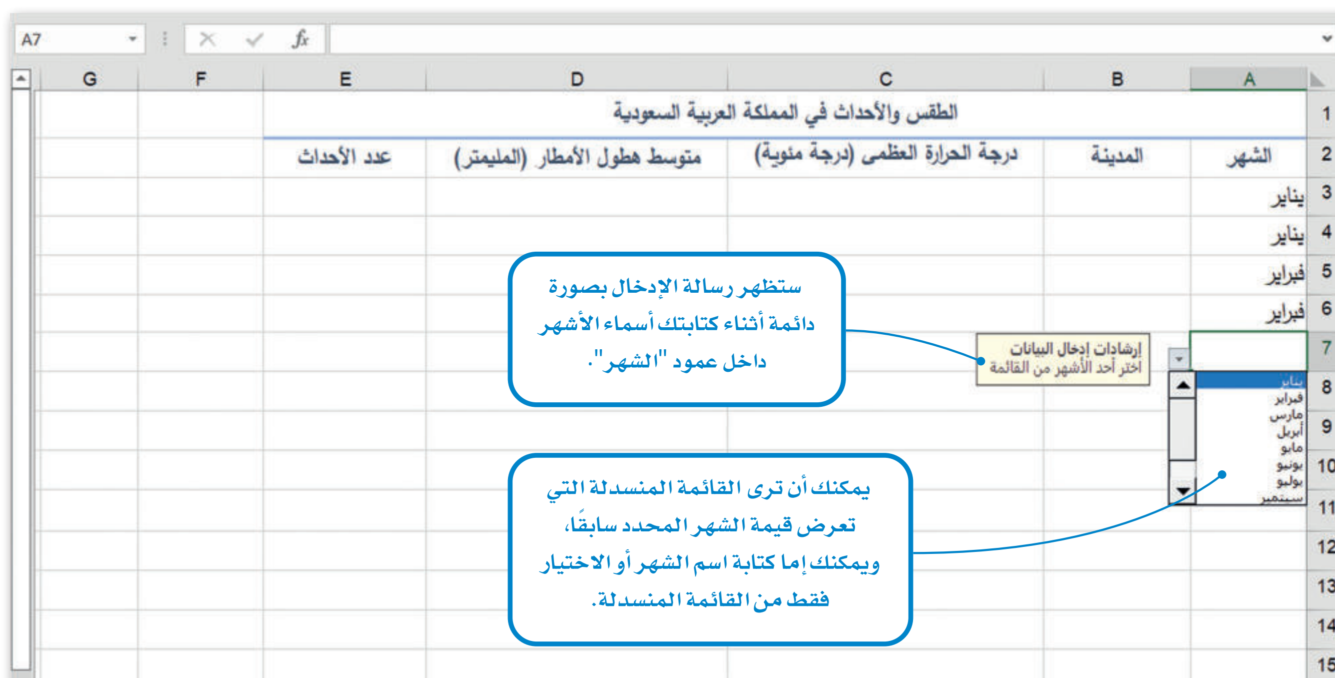
## لتعيين رسالة خطأ:

< من نافذة **تحقق من صحة البيانات** (Data Validation)، 1
اختر علامة التبويب **تنبيه إلى خطأ** (Error Alert). 2
< من مربع **النمط** (Style)، اختر **إيقاف** (Stop). 3
< في مربع **العنوان:** (Title:)، اكتب "المدخلات غير صالحة". 4
< في مربع **رسالة خطأ** (Error message) اكتب "يجب أن تختار أحد الأشهر من القائمة". 5
< اضغط على **موافق** (OK). 6

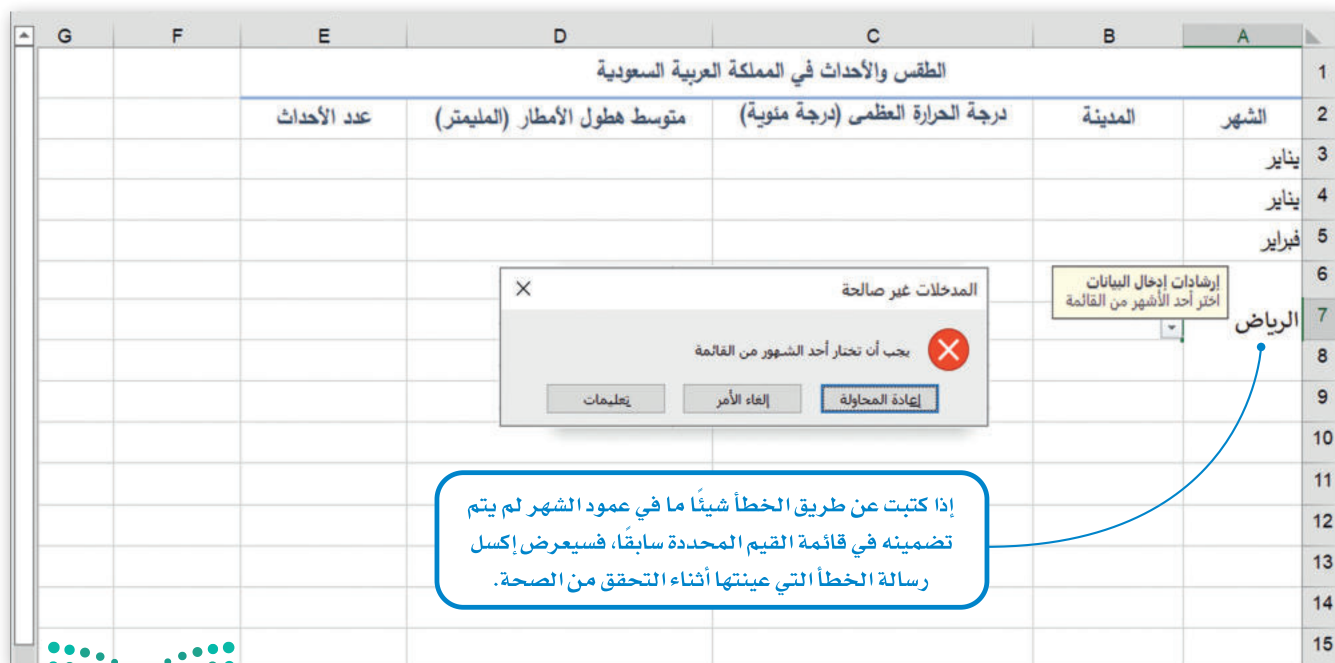



الشكل 2.24: تعيين رسالة تنبيه إلى الخطأ

بعد أن انتهيت من إجراء التحقق من صحة البحث والتحقق من صحة التواجد، يمكنك البدء في إدخال البيانات في الجدول. ففي كل خلية من **A3** إلى **A17** يمكنك رؤية قائمة منسدلة تعرض قيم أشهر محددة مسبقًا ويمكنك إما كتابة اسم الشهر أو الاختيار فقط من القائمة المنسدلة. عند إدخال قيم الأشهر سيتم عرض رسالة الإدخال "إرشادات إدخال البيانات"، أما إذا كتبت قيمة أخرى لم يتم تضمينها في قائمة القيم المحددة سابقًا في عمود الشهر، فستظهر رسالة الخطأ التي عينتها مسبقًا على الشاشة.




الشكل 2.25: قائمة الأشهر المنسدلة




الشكل 2.26: رسالة تنبيه إلى الخطأ الخاصة بالتحقق من صحة المدخلات

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | الطقس والأحداث في المملكة العربية السعودية | | | | |
| 2 | الشهر | المدينة | درجة الحرارة العظمى (درجة مئوية) | متوسط هطول الأمطار (المليمتر) | عدد الأحداث |
| 3 | يناير | | | | |
| 4 | يناير | | | | |
| 5 | فبراير | | | | |
| 6 | فبراير | | | | |
| 7 | مارس | | | | |
| 8 | مارس | | | | |
| 9 | أبريل | | | | |
| 10 | مايو | | | | |
| 11 | مايو | | | | |
| 12 | يوليو | | | | |
| 13 | يوليو | | | | |
| 14 | سبتمبر | | | | |
| 15 | أكتوبر | | | | |
| 16 | نوفمبر | | | | |
| 17 | نوفمبر | | | | |

بعد إدخال جميع أسماء الأشهر في الخلايا A3 إلى A17، سيظهر جدول "الطقس والأحداث في المملكة العربية السعودية" على هذا النحو في ورقة الإكسل.

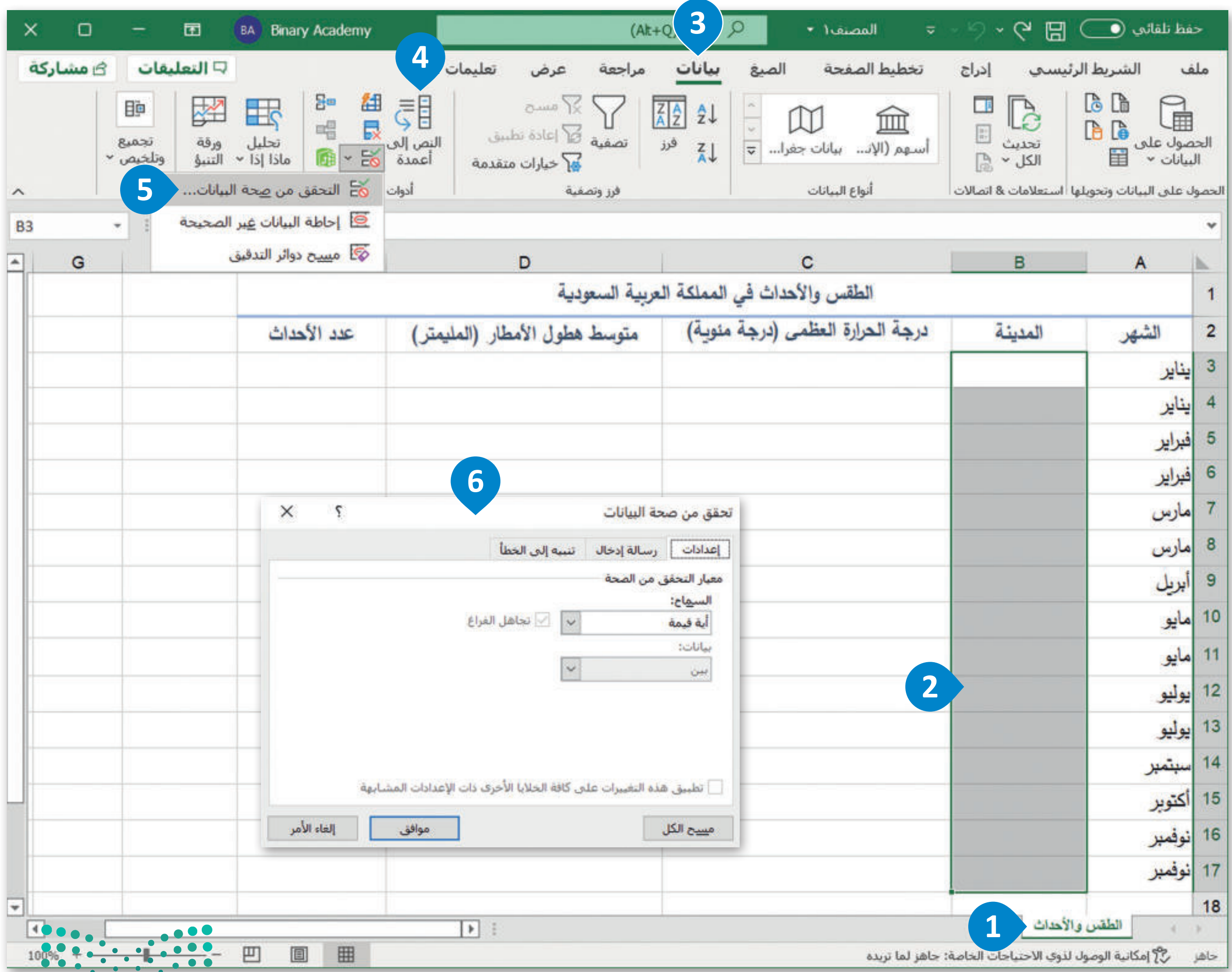
شكل 2.27: جدول التحقق من صحة البيانات

## التحقق من صحة الطول في إكسل Length Validation Check in Excel

ابدأ العمل في إدخال أسماء المدن في عمود "المدينة"، وقبل إدخال أسماء المدن تحقق من الطول؛ حتى تتمكن من إدخال قيم تتراوح بين 3 إلى 6 حروف فقط.

**لتطبيق التحقق من صحة الطول في إكسل:**

- انتقل إلى ورقة "**الطقس والأحداث في المملكة العربية السعودية**" في إكسل. 1
- حدد الخلايا من B3 إلى B17. 2
- ثم اختر علامة التبويب **بيانات** (Data). 3
- من مجموعة **أدوات البيانات** (Data tools)، 4 اضغط على نافذة **التحقق من صحة البيانات** (Data Validation). 5
- ستظهر نافذة **التحقق من صحة البيانات** (Data Validation). 6




الشكل 2.28: بدء عملية التحقق من صحة البيانات

### لتطبيق التحقق من صحة الطول في إكسل:

- من نافذة **تحقق من صحة البيانات** (Data Validation)، 1 اختر علامة التبويب **إعدادات** (Settings). 2
- من مربع **السماح** (Allow)، اختر **طول النص** (Text Length). 3
- من مربع **بيانات** (Data)، اختر **بين** (Between). 4
- في مربع **الحد الأدنى** (Minimum)، اكتب الرقم 3 وفي مربع **الحد الأقصى** (Maximum) اكتب الرقم 6. 5
- ألغ تحديد خيار **تجاهل الفراغ** (Ignore blank). 6

الشكل 2.29: تطبيق التحقق من صحة الطول في إكسل

### لتعيين رسالة إدخال غير صالح:

- من نافذة **تحقق من صحة البيانات** (Data Validation)، 1 اختر علامة التبويب **رسالة إدخال** (Input Message). 2
- في مربع **العنوان:** (Title:)، اكتب "إرشادات إدخال البيانات". 3
- في مربع **رسالة إدخال:** (Input message:)، اكتب "دخل اسم مدينة يتراوح بين 3 إلى 6 أحرف". 4

الشكل 2.30: تعيين رسالة إدخال غير صالح

### لتعيين رسالة تنبيه إلى الخطأ:

- من نافذة **تحقق من صحة البيانات** (Data Validation)، 1 اختر علامة التبويب **تنبيه إلى الخطأ** (Error Alert). 2
- من مربع **النمط** (Style)، اختر **إيقاف** (Stop). 3
- في مربع **العنوان** (Title:) اكتب "المدخلات غير صالحة". 4
- في مربع **رسالة خطأ** (Error message) اكتب "يجب أن يكون اسم المدينة يتراوح بين 3 إلى 6 أحرف". 5
- اضغط **موافق** (OK). 6

**نمط إشارة التحذير (Warning) لا يشجع على إدخال بيانات غير صالحة. رمز رسالة الخطأ مثلث أصفر مع علامة تعجب سوداء.**

**يعلن نمط المعلومات (Information) عن إدخال بيانات غير صالحة. رمز رسالة الخطأ هو فقاعة نص بيضاء اللون، بحرف "ا" صغير أزرق اللون.**

الشكل 2.31: تعيين رسالة تنبيه إلى الخطأ

بعد الانتهاء من إجراء التحقق من صحة الطول، يمكنك البدء في إدخال بيانات المدن، يُسمح لك في كل خلية من **B3** إلى **B17** كتابة اسم المدينة بحيث يتكون من 3 إلى 6 أحرف.

تظهر رسالة الإدخال باستمرار وإذا أدخلت عن طريق الخطأ قيمة أقل من 3 أحرف أو أكثر من 6 أحرف، فستظهر رسالة الخطأ التي قمت بتعيينها سابقًا على الشاشة.

ستظهر رسالة الإدخال دائمًا أثناء كتابة أسماء المدن في عمود "**المدينة**"

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | | | الطقس والأحداث في المملكة العربية السعودية | | | | |
| 2 | الشهر | المدينة | درجة الحرارة العظمى (درجة مئوية) | متوسط هطول الأمطار (المليمتر) | عدد الأحداث | | |
| 3 | يناير | جدة | | | | | |
| 4 | يناير | الرياض | | | | | |
| 5 | فبراير | الرياضالرياض | | | | | |
| 6 | فبراير | | | | | | |
| 7 | مارس | | | | | | |
| 8 | مارس | | | | | | |
| 9 | أبريل | | | | | | |
| 10 | مايو | | | | | | |
| 11 | مايو | | | | | | |
| 12 | يوليو | | | | | | |
| 13 | يوليو | | | | | | |
| 14 | سبتمبر | | | | | | |
| 15 | أكتوبر | | | | | | |
| 16 | نوفمبر | | | | | | |
| 17 | نوفمبر | | | | | | |
| 18 | | | | | | | |

إرشادات إدخال البيانات
أدخل اسم مدينة تتراوح بين 3 إلى 6 أحرف

المدخلات غير صالحة
يجب أن يكون اسم المدينة يتراوح بين 3 إلى 6 أحرف
إعادة المحاولة | إلغاء الأمر | تعليمات

الطقس والأحداث
جاهز | إمكانية الوصول لذوي الاحتياجات الخاصة: جاهز لما تريده | 100%

الشكل 2.32: رسالة الإدخال ورسالة تنبيه إلى الخطأ للتحقق من صحة الطول

إذا كتبت عن طريق الخطأ قيمة في عمود المدينة لا يتوافق مع المعايير التي حددتها بالفعل، فسيعرض إكسل رسالة الخطأ التي عينتها أثناء التحقق من الصحة.



الطقس والأحداث في المملكة العربية السعودية

| الشهر | المدينة | درجة الحرارة العظمى (درجة مئوية) | متوسط هطول الأمطار (المليمتر) | عدد الأحداث |
|---|---|---|---|---|
| يناير | جدة | | | |
| يناير | الرياض | | | |
| فبراير | جدة | | | |
| فبراير | الرياض | | | |
| مارس | جدة | | | |
| مارس | الرياض | | | |
| أبريل | الرياض | | | |
| مايو | جدة | | | |
| مايو | الرياض | | | |
| يوليو | جدة | | | |
| يوليو | جدة | | | |
| سبتمبر | الرياض | | | |
| أكتوبر | الرياض | | | |
| نوفمبر | جدة | | | |
| نوفمبر | الرياض | | | |

شكل 2.33: جدول التحقق من صحة البيانات

بعد استكمال ادخال جميع بيانات المدن في الخلايا B3 إلى B17، سيظهر جدول "الطقس والأحداث في المملكة العربية السعودية" على هذا النحو في ورقة الإكسل.



## التحقق من صحة النطاق في إكسل Range Validation Check in Excel

استمر في العمل على العمود الثالث، عمود "درجة الحرارة العظمى (درجة مئوية)". قبل إدخال قيم درجات الحرارة تحقق من صحة النطاق؛ حتى تتمكن من إدخال قيم درجات الحرارة بحيث تتراوح من 20 إلى 45 درجة مئوية فقط.

**لبدء عملية التحقق من صحة البيانات في إكسل:**

< انتقل إلى ورقة "الطقس والأحداث في المملكة العربية السعودية" في إكسل. 1

< حدد الخلايا من C3 إلى C17. 2

< ثم اختر علامة التبويب **بيانات** (Data). 3

< من مجموعة **أدوات البيانات** (Data tools)، اضغط على نافذة **التحقق من صحة البيانات** (Data Validation). 4 5

< ستظهر نافذة **التحقق من صحة البيانات** (Data Validation). 6

الشكل 2.34: بدء التحقق من صحة البيانات في إكسل



### لتطبيق التحقق من صحة النطاق في إكسل:

- من نافذة **تحقق من صحة البيانات** (Data Validation). 1
- اختر علامة التبويب **إعدادات** (Settings). 2
- من مربع **السماح** (Allow)، اختر **مخصص** (Custom). 3
- في مربع الصيغة، اكتب (AND(C3:C17>20;C3:C17<45. 4
- ألغ تحديد خيار **تجاهل الفراغ** (Ignore blank). 5

**تعني الصيغة "AND(C3:C17>20;C3:C17<45)" أن القيم التي ستدخلها في الخلايا من C3 إلى C17 يجب أن تكون أكبر من 20 درجة مئوية وأقل من 45 درجة مئوية.**




الشكل 2.35: تطبيق التحقق من صحة النطاق في إكسل

### لتعيين رسالة إدخال غير صالح:

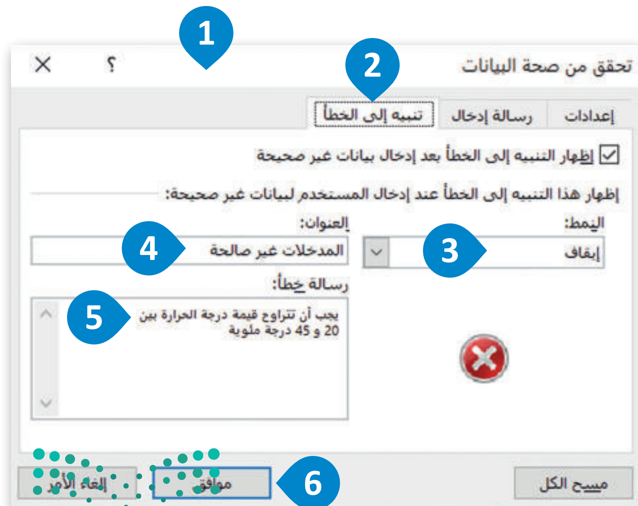
- من نافذة **تحقق من صحة البيانات** (Data Validation). 1
- اختر علامة التبويب **رسالة إدخال** (Input Message). 2
- في مربع العنوان: (:Title) اكتب "إرشادات إدخال البيانات". 3
- في مربع **رسالة إدخال:** (:Input message) اكتب "يجب أن تكون بيانات درجة الحرارة ضمن نطاق محدد من القيم". 4




الشكل 2.36: تعيين رسالة إدخال غير صالح

### لتعيين رسالة تنبيه إلى الخطأ:

- من نافذة **تحقق من صحة البيانات** (Data Validation). 1
- اختر علامة التبويب **تنبيه إلى الخطأ** (Error Alert). 2
- من مربع **النمط** (Style) اختر **إيقاف** (Stop). 3
- في مربع **العنوان:** (:Title) اكتب "المدخلات غير صالحة". 4
- في مربع **رسالة خطأ** (Error message) اكتب "يجب أن تتراوح قيمة درجة الحرارة بين 20 و45 درجة مئوية". 5
- اضغط **موافق** (OK). 6




الشكل 2.37: تعيين رسالة تنبيه إلى الخطأ

بعد الانتهاء من إجراء التحقق من صحة النطاق، يمكنك البدء في إدخال بيانات قيم درجات الحرارة العظمى (درجة مئوية).

يسمح لك في كل خلية من **C3** إلى **C17** إدخال قيمة درجة حرارة تتراوح بين 20 إلى 45 درجة مئوية.

ستظهر رسالة الإدخال دائمًا داخل الخلايا، وإذا كتبت عن طريق الخطأ قيمة أقل من 20 درجة مئوية أو أكبر من 45 درجة مئوية فستظهر رسالة الخطأ التي عينتها أثناء التحقق من الصحة على الشاشة.

عندما تضع رسالة في علامة تبويب "رسالة إدخال" ستظهر دائمًا أثناء كتابة قيم درجة الحرارة العظمى (درجة مئوية) في الخلايا من C3 إلى C17.

الطقس والأحداث في المملكة العربية السعودية

| الشهر | المدينة | درجة الحرارة العظمى (درجة مئوية) | متوسط هطول الأمطار (المليمتر) | عدد الأحداث |
|---|---|---|---|---|
| يناير | جدة | 28.80 | | |
| يناير | الرياض | 20.70 | | |
| فبراير | جدة | 29.80 | | |
| فبراير | الرياض | | | |
| مارس | جدة | 10 | | |
| مارس | الرياض | | | |
| أبريل | الرياض | | | |
| مايو | جدة | | | |
| مايو | الرياض | | | |
| يوليو | جدة | | | |
| يوليو | جدة | | | |
| سبتمبر | الرياض | | | |
| أكتوبر | الرياض | | | |
| نوفمبر | جدة | | | |
| نوفمبر | الرياض | | | |

إرشادات إدخال البيانات
يجب أن تكون بيانات درجة الحرارة ضمن نطاق محدد من القيم

المدخلات غير صالحة
يجب أن تتراوح قيمة درجة الحرارة بين 20 و 45 درجة مئوية
إعادة المحاولة | إلغاء الأمر | تعليمات

إذا كتبت عن طريق الخطأ قيمة في عمود درجة الحرارة العظمى (درجة مئوية) لا تفي بالمعايير التي حددتها بالفعل، فسيعرض إكسل رسالة الخطأ التي عينتها أثناء التحقق من الصحة في علامة تبويب "تنبيه إلى الخطأ".

الشكل 2.38: رسالة الإدخال ورسالة الخطأ للتحقق من صحة النطاق

| الشهر | المدينة | درجة الحرارة العظمى (درجة مئوية) | متوسط هطول الأمطار (المليمتر) | عدد الأحداث |
|---|---|---|---|---|
| يناير | جدة | 28.80 | | |
| يناير | الرياض | 20.70 | | |
| فبراير | جدة | 29.80 | | |
| فبراير | الرياض | 23.70 | | |
| مارس | جدة | 25.50 | | |
| مارس | الرياض | 28.00 | | |
| أبريل | الرياض | 33.60 | | |
| مايو | جدة | 30.70 | | |
| مايو | الرياض | 39.50 | | |
| يوليو | جدة | 38.20 | | |
| يوليو | جدة | 39.40 | | |
| سبتمبر | الرياض | 32.80 | | |
| أكتوبر | الرياض | 27.50 | | |
| نوفمبر | جدة | 27.60 | | |
| نوفمبر | الرياض | 20.40 | | |

**بعد استكمال إدخال جميع بيانات درجات الحرارة في الخلايا C3 إلى C17، سيظهر جدول "الطقس والأحداث في المملكة العربية السعودية" على هذا النحو في ورقة الإكسل.**

شكل 2.39: جدول التحقق من صحة البيانات



## التحقق من صحة الصيغة في إكسل Format Validation Check in Excel

استمر في العمل على العمود الرابع، عمود "متوسط هطول الأمطار (المليمتر)". قبل إدخال قيم هطول الأمطار، تحقق من صحة الصيغة؛ حتى تتمكن ليس فقط من إدخال القيم الصحيحة، ولكن أيضا الكسور العشرية. يتطلب هذا الإجراء تعيين الحد الأدنى والأقصى، لذلك ستعين الحد الأدنى لمتوسط قيمة هطول الأمطار يساوي 0 مم والحد الأقصى يساوي 30 مم.

**لبدء عملية التحقق من صحة البيانات في إكسل:**

- < انتقل إلى ورقة "الطقس والأحداث في المملكة العربية السعودية" في إكسل. 1
- < حدد الخلايا D3 إلى D17. 2
- < ثم اختر علامة التبويب بيانات (Data). 3
- < من مجموعة **أدوات البيانات** (Data tools)، 4 اضغط على نافذة **تحقق من صحة البيانات** (Data Validation). 5
- < ستظهر نافذة **تحقق من صحة البيانات** (Data Validation). 6

| الشهر | المدينة | درجة الحرارة العظمى (درجة مئوية) | متوسط هطول الأمطار (المليمتر) | عدد الأحداث |
|---|---|---|---|---|
| يناير | جدة | 28.80 | | |
| يناير | الرياض | 20.70 | | |
| فبراير | جدة | 29.80 | | |
| فبراير | الرياض | 23.70 | | |
| مارس | جدة | 25.50 | | |
| مارس | الرياض | 28.00 | | |
| أبريل | الرياض | 33.60 | | |
| مايو | جدة | 30.70 | | |
| مايو | الرياض | 39.50 | | |
| يوليو | جدة | 38.20 | | |
| يوليو | جدة | 39.40 | | |
| سبتمبر | الرياض | 32.80 | | |
| أكتوبر | الرياض | 27.50 | | |
| نوفمبر | جدة | 27.60 | | |
| نوفمبر | الرياض | 20.40 | | |

الشكل 2.40: بدء التحقق من صحة البيانات في إكسل

## لتطبيق التحقق من صحة الصيغة في إكسل:

< من نافذة **تحقق من صحة البيانات** (Data Validation)، 1 اختر علامة التبويب **إعدادات** (Settings). 2
< من مربع **السماح** (Allow) ، اختر **عدد عشري** (Decimal). 3
< من مربع **البيانات** (Data) اختر **بين** (Between). 4
< في مربع **الحد الأدنى** (Minimum)، اكتب 0، وفي مربع **الحد الأقصى** (Maximum): اكتب 30. 5
< ألغ تحديد خيار **تجاهل الفراغ** (Ignore Blank). 6

الشكل 2.41: تطبيق التحقق من صحة الصيغة في إكسل

## لتعيين رسالة إدخال غير صالح:

< من نافذة **تحقق من صحة البيانات** (Data Validation)، 1 اختر علامة التبويب **رسالة إدخال** (Input Message). 2
< في مربع **العنوان:** (Title:)، اكتب "إرشادات إدخال البيانات". 3
< في مربع **رسالة إدخال** (Input Message)، اكتب "يجب أن تكون قيمة هطول الأمطار رقم عشري." 4

الشكل 2.42: تعيين رسالة إدخال غير صالح

## لتعيين رسالة تنبيه إلى الخطأ:

< من نافذة **تحقق من صحة البيانات** (Data Validation)، 1 اختر علامة التبويب **تنبيه إلى الخطأ** (Error Alert). 2
< من مربع **النمط** (Style) اختر **إيقاف** (Stop). 3
< في مربع **العنوان:** (Title:) اكتب "المدخلات غير صالحة" . 4
< في مربع **رسالة خطأ** (Error message:)، اكتب "قيم هطول الأمطار ليست في شكل عشري." 5
< اضغط **موافق** (OK). 6

الشكل 2.43: تعيين رسالة تنبيه إلى الخطأ

عند الانتهاء من خطوة التحقق من صحة الصيغة، يمكنك البدء في إدخال بيانات هطول الأمطار. يمكن أن تدخل قيمة لكل خلية من **D3** إلى **D17** ويُسمح لك بإدخال قيمة هطول الأمطار في شكل عشري بقيم تتراوح بين 0 إلى 30 مم.

ستظهر رسالة الإدخال باستمرار وحال إدخال قيمة عن طريق الخطأ أقل من 0 مم أو أكبر من 30 ستظهر على الشاشة مباشرةً رسالة الخطأ التي قمت بتعيينها سابقًا.

عندما تضع رسالة في علامة تبويب "رسالة إدخال" ستظهر دائمًا أثناء كتابة متوسط قيم هطول الأمطار (مم) في الخلايا من D3 إلى D17.

الطقس والأحداث في المملكة العربية السعودية

| الشهر | المدينة | درجة الحرارة العظمى (درجة مئوية) | متوسط هطول الأمطار (المليمتر) | عدد الأحداث |
|---|---|---|---|---|
| يناير | جدة | 28.80 | 12.50 | |
| يناير | الرياض | 20.70 | 14.80 | |
| فبراير | جدة | 29.80 | | |
| فبراير | الرياض | 23.70 | -8 | |
| مارس | جدة | 25.50 | | |
| مارس | الرياض | 28.00 | | |
| أبريل | الرياض | 33.60 | | |
| مايو | جدة | 30.70 | | |
| مايو | الرياض | 39.50 | | |
| يوليو | جدة | 38.20 | | |
| يوليو | جدة | 39.40 | | |
| سبتمبر | الرياض | 32.80 | | |
| أكتوبر | الرياض | 27.50 | | |
| نوفمبر | جدة | 27.60 | | |
| نوفمبر | الرياض | 20.40 | | |

إرشادات إدخال البيانات
يجب أن تكون قيمة هطول الأمطار رقم عشري

المدخلات غير صالحة
قيم هطول الأمطار ليست في شكل عشري
إعادة المحاولة | إلغاء الأمر | تعليمات

الشكل 2.44: رسالة إدخال وتنبيه إلى الخطأ للتحقق من صحة النطاق

إذا كتبت عن طريق الخطأ قيمة في عمود متوسط هطول الأمطار (المليمتر) لا تفي بالمعايير التي حددتها بالفعل، فسيعرض إكسل رسالة الخطأ التي عينتها أثناء التحقق من الصحة.

الطقس والأحداث في المملكة العربية السعودية

| الشهر | المدينة | درجة الحرارة العظمى (درجة مئوية) | متوسط هطول الأمطار (المليمتر) | عدد الأحداث |
|---|---|---|---|---|
| يناير | جدة | 28.80 | 12.50 | |
| يناير | الرياض | 20.70 | 14.80 | |
| فبراير | جدة | 29.80 | 3.30 | |
| فبراير | الرياض | 23.70 | 8.30 | |
| مارس | جدة | 25.50 | 2.60 | |
| مارس | الرياض | 28.00 | 19.90 | |
| أبريل | الرياض | 33.60 | 23.70 | |
| مايو | جدة | 30.70 | 0.10 | |
| مايو | الرياض | 39.50 | 5.60 | |
| يوليو | جدة | 38.20 | 0.00 | |
| يوليو | جدة | 39.40 | 0.40 | |
| سبتمبر | الرياض | 32.80 | 0.00 | |
| أكتوبر | الرياض | 27.50 | 1.50 | |
| نوفمبر | جدة | 27.60 | 27.10 | |
| نوفمبر | الرياض | 20.40 | 20.00 | |

شكل 2.45: جدول التحقق من صحة البيانات

**بعد استكمال إدخال جميع بيانات متوسط هطول الأمطار في الخلايا D3 إلى D17، سيظهر جدول "الطقس والأحداث في المملكة العربية السعودية" على هذا النحو في ورقة الإكسل.**

## التحقق من صحة النوع في إكسل Type Validation Check in Excel

استمر في العمل على العمود الخامس، عمود "عدد الأحداث". قبل إدخال عدد الأحداث لكل مدينة، تحقق من صحة النوع؛ حتى تتجنب إدخال قيم سالبة. تستدعي هذه الخطوة تعيين قيمة الحد الأدنى، لذلك ستعين قيمة الحد الأدنى تساوي 1 حتى لا تكون قيم الحدث مساوية للصفر وكذلك لا تحوي قيمًا سالبة.

**لبدء عملية التحقق من صحة البيانات في إكسل:**

- انتقل إلى ورقة "الطقس والأحداث في المملكة العربية السعودية" في إكسل. 1
- حدد الخلايا من **E3** إلى **E17**. 2
- ثم اختر علامة تبويب **بيانات** (Data). 3
- من مجموعة **أدوات البيانات** (Data tools)، 4 اضغط على نافذة **تحقق من صحة البيانات** (Data Validation). 5
- ستظهر نافذة **تحقق من صحة البيانات** (Data Validation). 6

الشكل 2.46: بدء التحقق من صحة البيانات في إكسل

**لبدء عملية التحقق من صحة البيانات في إكسل:**

- من نافذة **تحقق من صحة البيانات** (Data Validation)، 1
- اختر علامة التبويب **إعدادات** (Settings). 2
- من مربع **السماح** (Allow) اختر **عدد صحيح** (Whole Number). 3
- من مربع **البيانات** (Data) اختر **أكبر من أو تساوي** (greater than or equal to). 4
- في مربع **الحد الأدنى:** (Minimum:) اكتب 1 5
- ألغ تحديد خيار **تجاهل الفراغ** (Ignore blank). 6

الشكل 2.47: تطبيق التحقق من صحة النوع في إكسل

**لتعيين رسالة إدخال غير صالح:**

- من نافذة **تحقق من صحة البيانات** (Data Validation)، 1
- اختر علامة التبويب **رسالة إدخال** (Input Message). 2
- في مربع **العنوان:** (Title:) اكتب "إرشادات إدخال البيانات". 3
- في مربع **رسالة إدخال** (Input Message)، اكتب أدخل رقمًا صحيحًا غير سالب. 4

الشكل 2.48: تعيين رسالة إدخال غير صالح

**لتعيين رسالة تنبيه إلى الخطأ:**

- من نافذة **تحقق من صحة البيانات** (Data Validation)، 1
- اختر علامة التبويب **تنبيه إلى الخطأ** (Error Alert). 2
- من مربع **النمط** (Style)، اختر **إيقاف** (Stop). 3
- في مربع **العنوان:** (Title:) اكتب "المدخلات غير صالحة". 4
- في مربع **رسالة خطأ** (Error message) اكتب "لا يمكن أن يكون عدد الأحداث سالبًا". 5
- اضغط **موافق** (OK). 6

الشكل 2.49: تعيين رسالة تنبيه إلى الخطأ

بعد الانتهاء من إجراء التحقق من صحة النوع، أدخل بيانات الأحداث كما هو موضح بالجدول رقم 2.8. يمكن أن تدخل قيمة لكل خلية من **E3** إلى **E17** ويُسمح لك بإدخال رقم يساوي أو أكبر من 1. سيتم عرض رسالة الإدخال باستمرار وإذا أدخلت عن طريق الخطأ قيمة أقل من 1، فستظهر على الشاشة رسالة الخطأ التي قمت بتعيينها سابقًا.

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | الطقس والأحداث في المملكة العربية السعودية | | | | |
| 2 | الشهر | المدينة | درجة الحرارة العظمى (درجة مئوية) | متوسط هطول الأمطار (المليمتر) | عدد الأحداث |
| 3 | يناير | جدة | 28.80 | 12.50 | 2 |
| 4 | يناير | الرياض | 20.70 | 14.80 | 5 |
| 5 | فبراير | جدة | 29.80 | 3.30 | 1 |
| 6 | فبراير | الرياض | 23.70 | 8.30 | |
| 7 | مارس | جدة | 25.50 | 2.60 | -5 |
| 8 | مارس | الرياض | 28.00 | 19.90 | |
| 9 | أبريل | الرياض | 33.60 | 23.70 | |
| 10 | مايو | جدة | 30.70 | | |
| 11 | مايو | الرياض | 39.50 | | |
| 12 | يوليو | جدة | 38.20 | | |
| 13 | يوليو | جدة | 39.40 | | |
| 14 | سبتمبر | الرياض | 32.80 | 0,00 | |
| 15 | أكتوبر | الرياض | 27.50 | 1,50 | |
| 16 | نوفمبر | جدة | 27.60 | 27,10 | |
| 17 | نوفمبر | الرياض | 20.40 | 20,00 | |

إرشادات إدخال البيانات
أدخل رقمًا صحيحًا غير سالب

المدخلات غير صالحة
لا يمكن أن يكون عدد الأحداث سالبًا
إعادة المحاولة | إلغاء الأمر | تعليمات

ستظهر رسالة الخطأ دائما أثناء كتابة عدد الأحداث.

إذا كتبت قيمة عن طريق الخطأ في عمود الأحداث لا تفي بالمعايير التي حددتها بالفعل، فسيعرض إكسل رسالة الخطأ التي عينتها أثناء التحقق من الصحة.

الشكل 2.50: رسالة إدخال وتنبيه إلى الخطأ للتحقق من صحة النوع



| الشهر | المدينة | درجة الحرارة العظمى (درجة مئوية) | متوسط هطول الأمطار (المليمتر) | عدد الأحداث |
|---|---|---|---|---|
| يناير | جدة | 28.80 | 12.50 | 2 |
| يناير | الرياض | 20.70 | 14.80 | 5 |
| فبراير | جدة | 29.80 | 3.30 | 1 |
| فبراير | الرياض | 23.70 | 8.30 | 8 |
| مارس | جدة | 25.50 | 2.60 | 1 |
| مارس | الرياض | 28.00 | 19.90 | 7 |
| أبريل | الرياض | 33.60 | 23.70 | 1 |
| مايو | جدة | 30.70 | 0.10 | 1 |
| مايو | الرياض | 39.50 | 5.60 | 1 |
| يوليو | جدة | 38.20 | 0,00 | 1 |
| يوليو | جدة | 39.40 | 0.40 | 2 |
| سبتمبر | الرياض | 32.80 | 0.00 | 4 |
| أكتوبر | الرياض | 27.50 | 1.50 | 4 |
| نوفمبر | جدة | 27.60 | 27.10 | 1 |
| نوفمبر | الرياض | 20.40 | 20.00 | 5 |

**بعد استكمال ادخال الأحداث في الخلايا من E3 إلى E17، سيظهر جدول "الطقس والأحداث في المملكة العربية السعودية" على هذا النحو في ورقة الإكسل.**

شكل 2.51: جدول التحقق من صحة البيانات



# تمرينات

1

| حدد الجملة الصحيحة والجملة الخاطئة فيما يلي: | صحيحة | خاطئة |
|---|---|---|
| 1. يشير التحقق من صحة البيانات إلى الإجراء الذي يحذف تلقائيًا أي بيانات أولية لا تفي بمعايير معينة. | ○ | ○ |
| 2. لا يوجد سوى خمسة أنواع للتحقق من صحة البيانات. | ○ | ○ |
| 3. يساعد التحقق من التواجد على تقليل الأخطاء باستخدام قائمة محدودة من القيم المحددة مسبقًا. | ○ | ○ |
| 4. يهدف التحقق من البحث إلى التأكد من أن الرموز والحروف تدخل بنطاق طول محدد. | ○ | ○ |
| 5. يُستخدم فحص النطاق للتأكد من أن الأرقام التي تُدخل تقع ضمن نطاق معين. | ○ | ○ |
| 6. يُستخدم التحقق من الصيغة للتأكد من أن البيانات تأتي بصيغة محددة مسبقًا. | ○ | ○ |
| 7. يساعد التحقق من النوع في تقليل أخطاء اللغة. | ○ | ○ |
| 8. يتم استخدام رقم التحقق إذا أردت التأكد من إدخال مجموعة من الأرقام بشكل صحيح. | ○ | ○ |
| 9. يعتبر مايكروسوفت إكسل الأداة الوحيدة التي يمكن استخدامها للتحقق من صحة البيانات. | ○ | ○ |
| 10. يمكن إجراء التحقق من صحة البيانات بعد إدخال القيم في برنامج التحقق من صحة البيانات. | ○ | ○ |



2 اشرح بإيجاز المقصود بعملية التحقق من صحة البيانات.

3 صِف باختصار الخطوات التي يجب اتباعها في إكسل لإكمال عملية التحقق من صحة البيانات.



**4** أنشئ دليل عناوين لمعلومات أصدقائك يتضمن الحقول التالية: الاسم، والهاتف، وعنوان المنزل، وعنوان البريد الإلكتروني، وتاريخ الميلاد، والهواية. اكتب نوع التحقق من صحة البيانات الذي يجب إجراؤه على كل حقل من الحقول السابقة.

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

**5** قارن ما يلي: (أ) فحص الطول مقابل فحص النطاق، (ب) فحص الصيغة مقابل فحص النوع. أعطِ أمثلة على استخدام كل نوع من أنواع التحقق السابقة.

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________



# المشروع

1

افترض أنك تعمل كباحث في القطاع الصحي وتريد إعداد تقرير حول مشكلة مرض السكري في بلدك، اشرح الخطوات التي تمكنك من التحقق من صحة البيانات التي قد تم جمعها.

2

على وجه التحديد يتعين عليك الإجابة على أسئلة مثل الأسئلة التالية:

ما نوع الأعمدة التي ستقوم بإنشائها؟

أيٍ من عمليات التحقق الستة سيتم إجراؤها في كل عمود؟ ولماذا؟



3

أنشئ عرضًا تقديميًا في باوربوينت (PowerPoint)، موضحًا خطوات إجراء التحقق من صحة البيانات التي قمت بها؟



# ماذا تعلمت

- ما المقصود بجمع البيانات.
- التمييز بين أنواع البيانات المختلفة.
- كيفية ترميز البيانات.
- كيفية التحقق من صحة البيانات.
- كيفية التحقق من جودة المعلومات.
- كيفية التمييز بين مصادر البيانات الرئيسة ومصادر البيانات الثانوية.

## المصطلحات الرئيسة

| | | | |
|---|---|---|---|
| جودة المعلومات | Information Quality | الدقة | Accuracy |
| رقم الكتاب المعياري الدولي | Isbn | الملاءمة | Appropriateness |
| الأذونات القانونية | Legal Permissions | البيانات الصوتية | Audio Data |
| التحقق من الطول | Length Check | الرموز الشريطية | Barcode |
| مستوى التفاصيل | Level Of Detail | التحقق من الرقم | Check Digit |
| التحقق من البحث | Lookup Check | الكفاية | Completeness |
| منصات البيانات المفتوحة | Open Data Platforms | البيانات | Data |
| التحقق من التواجد | Presence Check | ترميز البيانات | Data Coding |
| رمز الاستجابة السريعة | QR Code | أنواع البيانات | Data Types |
| التحقق من النطاق | Range Check | التحقق من صحة البيانات | Data Validation |
| التحقق من النوع | Type Check | بيانات ثابتة | Fixed Data |
| بيانات متغيرة | Variable Data | التحقق من الصيغة | Format Check |
| بيانات الفيديو | Video Data | البيانات المرئية | Graphical Data |

# 3. التحليل الاستكشافي للبيانات



في الوحدات السابقة، تعلم الطالب مفهوم البيانات وأنماطها المختلفة وكيفية جمعها بشكل صحيح. وفي هذه الوحدة، سيتعلم الطالب طريقة فحص وتحليل البيانات لفهمها بشكل أفضل.

## أهداف التعلم

بنهاية هذه الوحدة سيكون الطالب قادرًا على أن:

- يصنِّف أنماط تحليل البيانات.
- يعرِّف مفهوم التحليل الاستكشافي للبيانات.
- يصنِّف أنماط التحليل الاستكشافي للبيانات.
- يشرح مراحل عملية التحليل الاستكشافي للبيانات.
- يعرِّف مفهوم المكتبات البرمجية.
- يطوِّر برنامج لتحليل البيانات باستخدام مكتبات البرمجة.
- يستخدم تقنيات إعداد البيانات وتنظيفها في مجموعة بيانات.
- يشرح أهمية تصوير البيانات.
- يميِّز الأنماط المختلفة من الرسوم البيانية التي يمكن إنشاؤها باستخدام مكتبات البايثون.




رابط الدرس الرقمي
www.ien.edu.sa

الدرس الأول

# تحليل البيانات

## مفهوم تحليل البيانات What is Data Analysis

أنت تحلل كل شيء في حياتك اليومية، فعلى سبيل المثال، عند التفكير فيما حدث في المرة الأخيرة التي فعلت فيها شيئًا وماذا سيحدث إذا اتخذت هذا القرار مرة أخرى. وما هذا إلاّ تحليل لماضيك أو لمستقبلك واتخاذ قراراتك بناءًا على ذلك التحليل.

يُعرّف تحليل البيانات بأنه عملية فحص البيانات وتنظيفها وتحويلها ونمذجتها لاستخلاص المعلومات المفيدة، وإثراء الاستنتاجات ودعم عملية اتخاذ القرارات.

**تحليل البيانات:**

فحص منهجي للبيانات عن طريق العينات والقياس والتصوير.

## أنماط تحليل البيانات Types of Data Analysis

بناءًا على رغبتك في تحليل البيانات والمشكلة المحددة التي تحاول حلها، فقد تقوم بإجراء أنواع مختلفة من التحليلات:

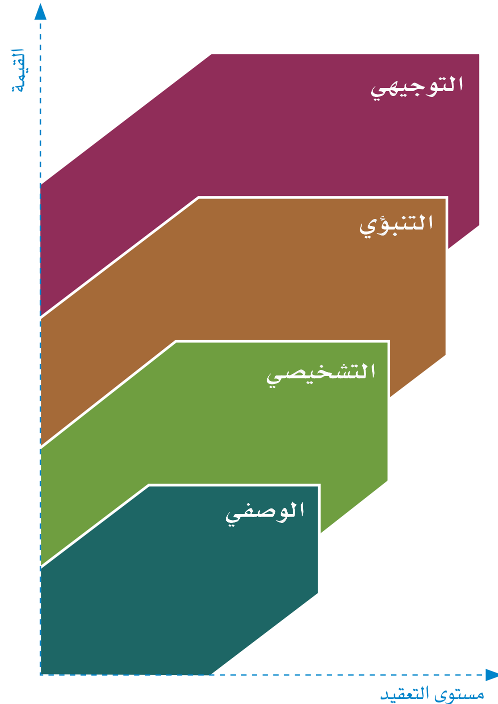
- **التحليل التوجيهي** (Prescriptive Analysis).
- **التحليل التنبؤي** (Predictive Analysis).
- **التحليل التشخيصي** (Diagnostic Analysis).
- **التحليل الوصفي** (Descriptive Analysis).

## التحليل الوصفي Descriptive Analysis

يهتم التحليل الوصفي بما حدث في الماضي، ويُعرف غالبًا بإسم التحليلات الوصفية أو الإحصاءات الوصفية وهو وصف أو تلخيص مجموعة من البيانات باستخدام التقنيات الإحصائية. وتبرُز أهمية التحليل الوصفي كأحد أهم أنماط تحليل البيانات، نظرا لقدرته على تحليل البيانات التي لم يتم تحليلها بعد. إلا أن التحليلات الوصفية لا تُقدّم تنبؤات حول المستقبل.

## التحليل التشخيصي Diagnostic Analysis

يهتم تحليل البيانات التشخيصي بأسباب حدوث شيء ما. ويقوم التحليل التشخيصي على التحليل الوصفي، ويسعى المختصون من خلاله إلى فهم سبب التوجهات والأنماط التي تمت ملاحظتها.




الشكل 3.1: أنماط تحليل البيانات

## التحليل التنبؤي Predictive Analysis

يهتم تحليل البيانات التنبؤي بمحاولة التنبؤ بالنتائج المستقبلية بناءً على الاتجاهات المكتشفة سابقًا والبيانات التاريخية، باستخدام تقنيات النمذجة والإحصاءات. ويستخدم التحليل التنبؤي في العديد من الحالات المختلفة، مثل التنبؤ بالطقس وسياسات التأمين وغير ذلك.

**التحليل التنبؤي:**
استخدام بيانات سابقة مع نماذج رياضية للتنبؤ بالنتائج المستقبلية أو الأحداث غير المعروفة.

## التحليل التوجيهي Prescriptive Analysis

وتتمثل المرحلة الأخيرة من تحليل البيانات في التحليل التوجيهي الذي يهتم بمحاولة إيجاد مسار العمل الأمثل. بناءً على اكتشافات مراحل التحليل السابقة، فإن الهدف من التحليلات التوجيهية هو تقديم توصيات للخطوات المستقبلية. وهذا النوع من التحليل مفيد بشكل خاص في القطاع الصحي، وذلك للحاجة إلى توصيات طبية آمنة.

**التحليلات التنبؤية والتوجيهية أكثر تعقيدًا من التحليلات الوصفية والتشخيصية، لكنها توفر المزيد من القيمة المضافة والفهم الأعمق لكل مشروع.**

## عملية تحليل البيانات Data Analysis Process

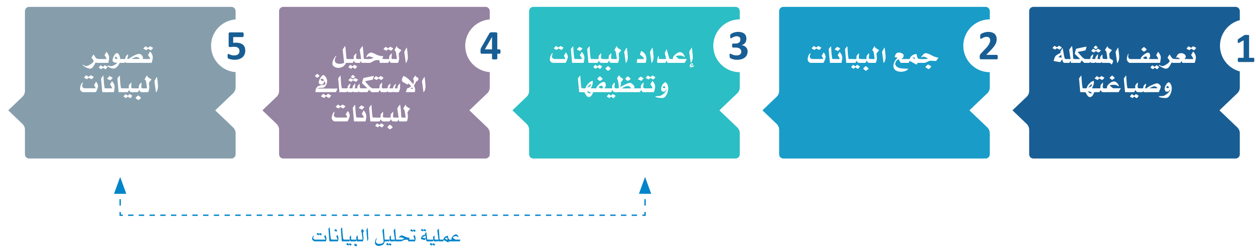
تتضمن عملية تحليل البيانات جمع المعلومات ومعالجتها وفحص البيانات. وبناءً على نتائج المعالجة، يمكنك اتخاذ القرارات أو استخلاص النتائج.

خطوات عملية تحليل البيانات هي كالآتي:

> **إعداد البيانات وتنظيفها:** من خلال هذه العملية تزيل المسافات البيضاء الزائدة والسجلات المكررة وأخطاء البيانات الأساسية. فتنظيف البيانات إلزامي قبل إرسال المعلومات لتحليلها.

> **التحليل الاستكشافي للبيانات:** وفي هذه الخطوة تبدأ في استخدام برامج تحليل البيانات وأدوات أخرى لمساعدتك على تفسير البيانات وفهمها والتوصل إلى استنتاجات.

> **تصوير البيانات:** وفي هذه الخطوة تستطيع أن تقوم بتصوير البيانات من خلال الرسم البياني لمجموعة من المعلومات أو البيانات. ويُسهل تصوير البيانات عملية فهمها وتحليلها. فباستخدام العناصر المرئية مثل المخططات والرسوم البيانية والخرائط، يُصبح تصور البيانات أكثر سهولة وفهمًا وقابلية للاستخدام.




الشكل 3.2: دورة حياة علم البيانات



## تعريف مفهوم التحليل الاستكشافي للبيانات
## What is Exploratory Data Analysis

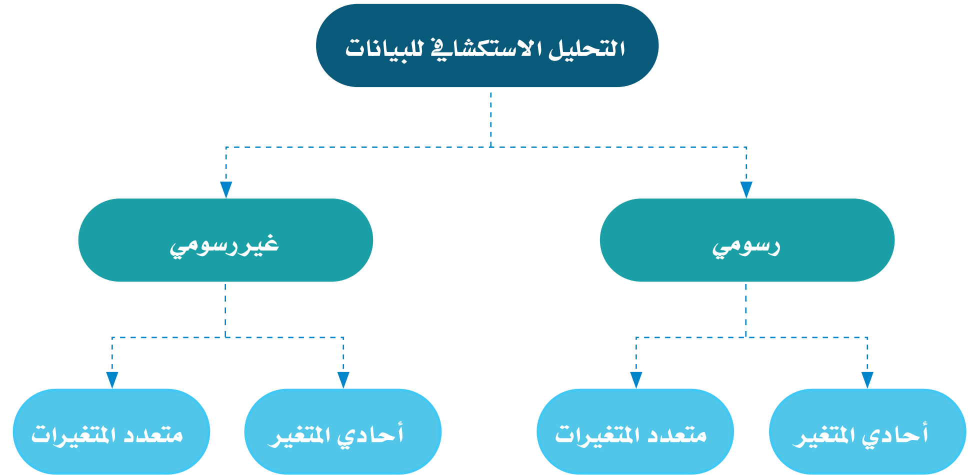
بشكل عام، يجدر بك محاولة فهم البيانات وجمع أكبر قدر ممكن من المعلومات قبل الانتقال إلى مهمة النمذجة. التحليل الاستكشافي للبيانات هو وسيلة لفهم البيانات وإجراء الفحوصات الأولية وتلخيص خصائصها الرئيسية. وتكمن الأهداف الرئيسية للتحليل الاستكشافي للبيانات في اكتشاف الاتجاهات والأنماط والميزات الجديدة في البيانات. يمكنك أيضًا اكتشاف الحالات الشاذة في مجموعة البيانات، واختبار فرضيتك الأولية والحصول على فهم أفضل لمتغيرات مجموعة البيانات والعلاقات بينها. يمكن أن يساعدك التحليل الاستكشافي للبيانات أيضًا في تحديد الأخطاء الواضحة والتأكد من أن نتائج مهمة معينة صالحة وقابلة للتطبيق على أي هدف مرغوب. ونظرًا لأن استخلاص المعلومات من خلال النظر إلى الأرقام يمكن أن يكون عملية مملة، فقد تم تطوير التحليل الاستكشافي للبيانات كمساعدة للقيام بهذه العملية، و يجري تحقيق كل ذلك بمساعدة الموجز الإحصائي والتمثيلات الرسومية وطرق تصوير البيانات. وبمجرد اكتمال التحليل الاستكشافي للبيانات واستخلاص معلومات كافية من البيانات، فيمكنك استخدام هذه الخصائص لإجراء أساليب تحليل أكثر تعقيدًا مثل تعلّم الآلة.

**تحليل البيانات الاستكشافي:**

المنهج المتبع في تحليل مجموعة من البيانات لتلخيص خصائصها الرئيسة، وغالبا تتم باستخدام الطرق البصرية.

## أنماط التحليل الاستكشافي للبيانات
## Types of Exploratory Data Analysis

يُصنف التحليل الاستكشافي للبيانات بطريقتين: أولًا، طريقة رسومية أو غير رسومية، وثانيًا، طريقة أحادية أو متعددة المتغيرات (عادةً ما يكون المتغير ثنائيًا). في حالة أحادي المتغير فإنه يتم تحليل تأثير متغير مستقل واحد فقط، بينما مع متعدد المتغيرات يتم تحليل أكثر من متغير مستقل في آن واحد، وفي المشاريع الكبيرة يُعد التحليل متعدد المتغيرات الأكثر شيوعا.




الشكل 3.3: أنماط التحليل الاستكشافي للبيانات



## التحليل غير الرسومي Non-Graphical Analysis

### التحليل غير الرسومي أحادي المتغير Univariate Non-Graphical Analysis

أحد الأمثلة على التحليل غير الرسومي أحادي المتغير هو تأثير العمر على الإصابة ببعض أنواع الأمراض مثل مرض الزهايمر. هذا التحليل أحادي المتغير لأنه يقيس تأثير العمر فقط، وغير رسومي لأنه لا يستخدم تقنيات التصوير البياني.

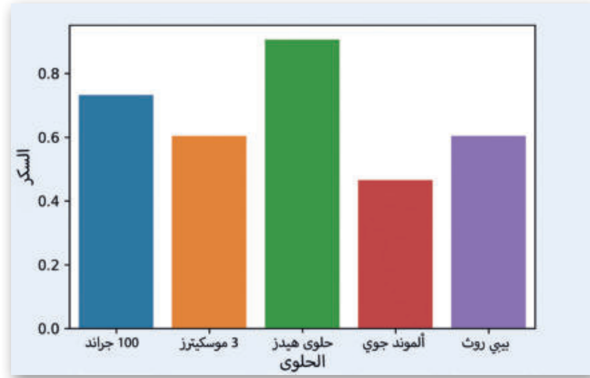
### التحليل غير الرسومي متعدد المتغيرات Multivariate Non-Graphical Analysis

إذا أخذت في الاعتبار من المثال السابق آثار النظام الغذائي والتمارين الذهنية والوراثة أيضًا، فسيكون هذا التحليل تحليلًا غير رسومي متعدد المتغيرات.

## التحليل الرسومي Graphical Analysis

### التحليل الرسومي أحادي المتغير Univariate Graphical Analysis

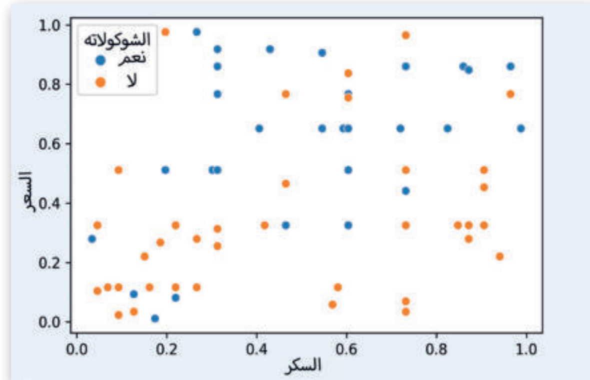
في الشكل 3.4 مثال على التحليل الرسومي أحادي المتغير. الشكل يُظهر **مخطط شريطي** (Bar Chart)، ويمثل كل عمود نسبة السكر التي تحتويها خمس أنواع من الحلوى المختلفة، وهذا التحليل الرسومي أحادي المتغير لأنه يحلل متغير واحد فقط، ويتم عرضه بطريقة رسومية.



الشكل 3.4: التحليل الرسومي أحادي المتغير

### التحليل الرسومي متعدد المتغيرات Multivariate Graphical Analysis

في الشكل 3.5 مثال على التحليل الرسومي متعدد المتغيرات. الشكل يُظهر **مخطط نقطي** (Scatter Plot) يعبّر عن أنواع الحلوى المختلفة بحيث يمثل المحور الأفقي (س) محتوى السكر في الحلوى بينما المحور الرأسي (ص) سعر الحلوى، ويتم ترميزها بالألوان بناءً على ما إذا كانت الحلوى تحتوي على الشوكولاتة أم لا. ستتعرف على المخطط النقطي وأنواع أخرى من أشكال تصوير البيانات لاحقًا في هذه الوحدة، وهذا المثال على التحليل الرسومي متعدد المتغيرات تم فيه تحليل ثلاث متغيرات معًا، وتصوير علاقة الارتباط بينهم بشكل بياني.



الشكل 3.5: التحليل الرسومي متعدد المتغيرات

## أدوات تحليل البيانات Data Analysis Tools

هناك العديد من الأدوات التي يمكنك استخدامها لمعالجة العلاقات والارتباطات بين مجموعات البيانات ومعالجتها وتحليلها، وتساعدك هذه الأدوات أيضًا في تحديد الأنماط والتوجهات للتفسير. لاختيار أداة تحليل البيانات، يجب عليك أولاً فهم احتياجاتك. الأداة التحليلية الأكثر شيوعًا واستخدامًا على نطاق واسع في جميع الصناعات تقريبًا هي إكسل. بالإضافة إلى برامج جداول البيانات، يمكن إجراء تحليل البيانات بلغات البرمجة المتخصصة والبيئات البرمجية أيضًا. البيئات الأكثر شعبية هي: **مفكرة جوبيتر** (Jupyter Notebook)، و **آر ستوديو** (RStudio)، و **ماتلاب** (MATLAB).

في هذه الوحدة، ستستخدم مفكرة جوبيتر كأداة لتحليل البيانات.

## تحليل البيانات باستخدام البايثون Data Analysis with Python

كما تعلمت سابقًا، يمكن استخدام البايثون في تحليل البيانات، وهي واحدة من اللغات الأكثر استخدامًا لمشروعات علم البيانات من قبل كل من علماء البيانات ومطوري البرامج. يمكن استخدامه للتنبؤ بالنتائج، وأتمتة الوظائف، وتبسيط العمليات، وتمكين ذكاء الأعمال. لإجراء تحليل البيانات باستخدام لغة البايثون، يمكنك الاستفادة من مكتبات البايثون.

## المكتبات والوحدات النمطية التابعة لبايثون Python Libraries/Modules

عادة ما تكون المكتبة عبارة عن مجموعة من الكتب أو موقع يتم فيه الاحتفاظ بالعديد من الكتب لاستخدامها لاحقًا، ولكن في البرمجة، المكتبة عبارة عن مجموعة من المقاطع البرمجية المكتوبة مسبقًا والبرامج الفرعية التي يمكن أن يستخدمها البرنامج، وصُمّمت لمساعدة كل من المبرمج و**المحول البرمجي** (Compiler) على إنشاء برنامج. ولاستخدام المكتبة، عليك تضمينها في مقطعك البرمجي، ولاستخدامها في البايثون، عليك إضافة كلمة **استيراد** (Import) واسم المكتبة.

المكتبة في لغات البرمجة مثل البايثون هي مجموعة من المقاطع البرمجية المجهزة مسبقًا والتي يمكن استخدامها لاحقًا في البرنامج لعمليات محددة جيدًا. فبالمقارنة مع لغات البرمجة الأخرى، لا تتعلق المكتبة بأي سياق محدد في البايثون، وقد تحتوي المكتبة أيضًا على **التوثيق** (Documentation) **وبيانات التكوين** (Configuration Data) **وقوالب الرسائل** (Message Templates) **والفئات** (Classes) **والقيم** (Values) وغيرها.

في البايثون، تصف المكتبة بشكل عام مجموعة من الوحدات النمطية، وتحوي أساسًا مجموعة من الوحدات النمطية أو حزم من المقاطع البرمجية القابلة لإعادة الاستخدام في العديد من البرامج، فهي تُبسّط وتُسرّع برمجة البايثون للمطورين لأنهم غير مضطرين لإعادة كتابة نفس المقطع البرمجي لبرامج مختلفة. يعتمد تعلّم الآلة وعلوم البيانات وتصوير البيانات والمجالات الأخرى بشكل كبير على مكتبات البايثون.

**الجدول 3.1: مزايا وعيوب استخدام مكتبات المقاطع البرمجية**

| + الإيجابيات | - السلبيات |
|---|---|
| سرعة الإعداد والاستخدام في مقطعك البرمجي. | إذا كنت بحاجة إلى تغييرات، فمن الصعب جدًا أو المستحيل تنفيذها. |
| عادةً خالية من الأخطاء وتعمل كما هو متوقع. فلا يلزم تصحيح الأخطاء والاختبار. | لا تعلم إذا كان سيستمر دعم المكتبة لنفس المدة الزمنية التي سيتم بها استخدام مقطعك البرمجي. |
| عادةً ما تكون محسَّنة وسريعة. | |
| لا حاجة لتعلم خوارزميات معقدة لتنفيذها. | |



## مكتبة البايثون القياسية Python Standard Library

مكتبة البايثون القياسية هي مجموعة دقيقة من تراكيب اللغة ووحداتها ودلالتها، وتم تضمينها في توزيع البايثون القياسي، وتتعامل المكتبة مع وظائف أساسية مختلفة مثل **المدخلات والمخرجات** (Input/Output- I/O) والوحدات الأساسية الأخرى. وتحظى لغة البايثون بشعبية كبيرة بسبب كل هذه الوظائف، حيث تضم المكتبة القياسية أكثر من 200 وحدة نمطية.

## مكتبات البايثون لعلم البيانات Python Libraries for Data Science

على الرغم من أنه يمكنك التعامل مع البيانات في البايثون؛ إلا أن هناك العديد من المكتبات مفتوحة المصدر التي تجعل مشروعات علوم البيانات أسهل بكثير. ويرد في الجدول التالي بعض المكتبات المستخدمة لكل مهمة من مهام علم البيانات.

**الجدول 3.2: مكتبات البايثون لعلم البيانات**

| مهام علم البيانات | المكتبات |
|---|---|
| تنقيب البيانات | Scrapy, Beautiful Soup, Requests |
| معالجة البيانات / الحوسبة العلمية | NumPy, SciPy, pandas, TensorFlow, Keras, scikit-learn, PyBrain, PyTorch, OpenCV, Mahotas |
| تصوير البيانات | Matplotlib, seaborn, Altair, Bokeh, plotly |

ستستخدم في هذه الوحدة:

> مكتبة **نمباي** (NumPy)، للعمليات الحسابية والرياضية.

> مكتبة **باندس** (Pandas)، للتعامل مع البيانات والتعديل عليها بطرق مختلفة.

> مكتبة **مات بلوت ليب** (Matplotlib)، لتصوير البيانات.

**مفكرة جوبيتر ليست بيئة تطوير متكاملة (IDE) لبايثون ولكن تم تحسينها لمشروعات علوم البيانات.**

## مفكرة جوبيتر Jupyter Notebook




في هذه الوحدة، ستستخدم مفكرة جوبيتر كأداة لتحليل البيانات، وهي تطبيق إلكتروني على الإنترنت لإنشاء المستندات الحسابية ومشاركتها. كل مستند يسمى مفكرة ويتضمن مقاطعك البرمجية وتعليقاتك وبياناتك الأولية والمعالجة والبيانات المصورة، ويمكنك أيضًا تخزين البيانات في ملف خارجي أو دمجها في المفكرة. ومن الجدير بالذكر أن بيئة جوبيتر تدعم لغات برمجة مختلفة بما فيها البايثون، علاوة على ذلك، يمكنك من خلالها إنشاء مخرجات تفاعلية مثل HTML أو مقاطع الفيديو.



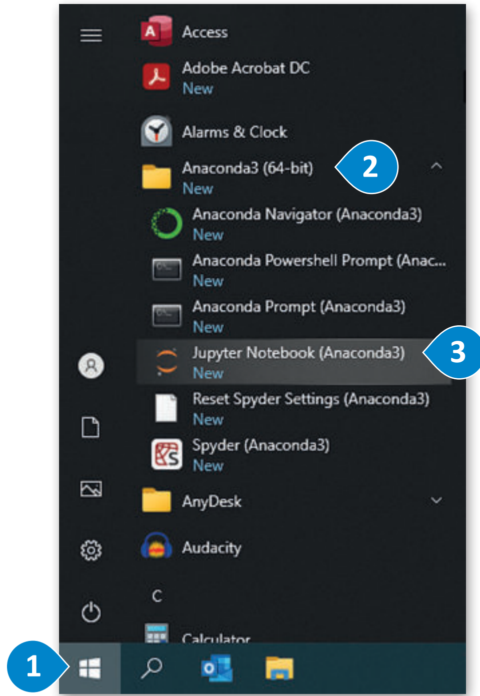

في هذه الوحدة، ستستخدم الإصدار **غير المتصل بالإنترنت** (Offline) من مفكرة جوبيتر، وأسهل طريقة لتثبيتها محليًا هي من خلال **اناكوندا** (Anaconda)، وهي منصة توزيع مفتوحة المصدر للطلبة والهواة. يمكنك تنزيل وتثبيت اناكوندا من هنا: https://www.anaconda.com/products/distribution، وسيتم تثبيت البايثون و مفكرة جوبيتر تلقائيًا.

**لفتح مفكرة جوبيتر:**

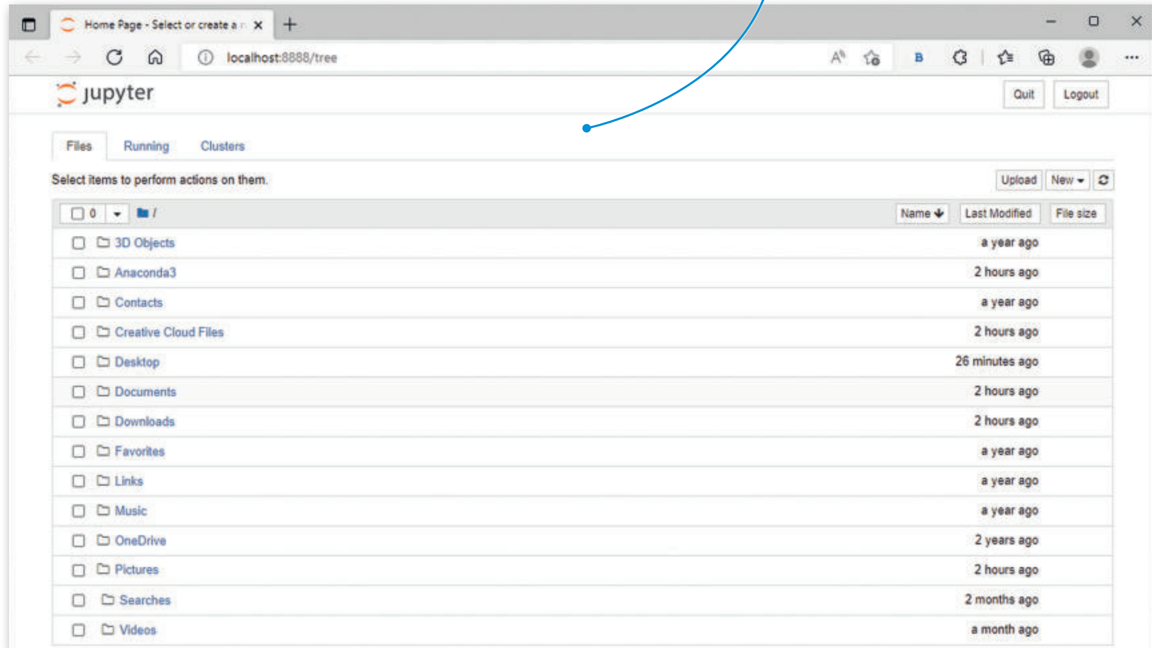
< اضغط Start (ابدأ). 1 ثم اضغط Anacoda3 (اناكوندا3). 2

< اختر Jupyter Notebook (مفكرة جوبيتر). 3

< ستفتح صفحة Jupyter Notebook (مفكرة جوبيتر) الرئيسية في المتصفح.




الصفحة الرئيسية لمفكرة جوبيتر.




الشكل 3.6: الصفحة الرئيسية لمفكرة جوبيتر

## نبذة تاريخية

عرّف عالم الرياضيات الأمريكي **جون توكي** (John Tukey) تحليل البيانات في عام 1961 على النحو التالي: "هي إجراءات لتحليل البيانات، وتقنيات لتفسير نتائج هذه الإجراءات، وطرق لتخطيط جمع البيانات لجعل تحليلها أسهل وأكثر دقة، وفي المجمل هي كل الآليات ونتائج الإحصاءات (الرياضية) ذات العلاقة بتحليل البيانات."

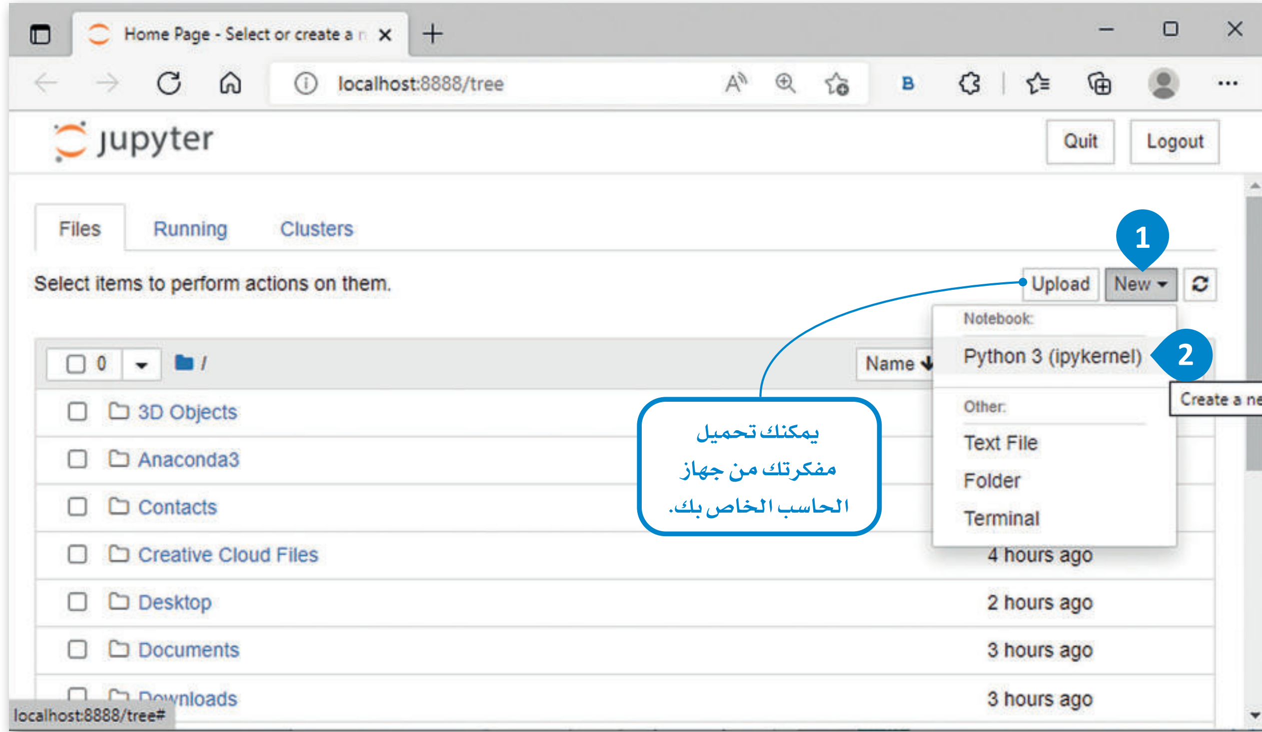


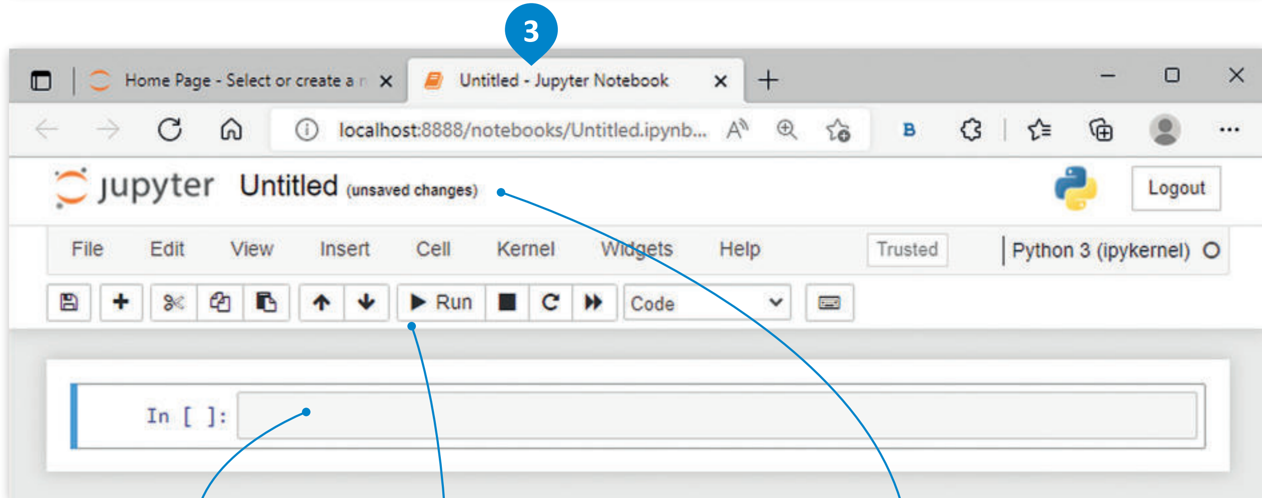
### لإنشاء مفكرة جديدة في جوبيتر:

< في الزاوية اليمنى العلوية من شاشتك، اضغط على New (جديد). 1

< حدد Python 3 (Ipykernel) (البايثون 3). 2

< وستفتح المفكرة الخاصة بك في علامة تبويب جديدة في المتصفح الخاص بك. 3




الشكل 3.7: إنشاء مفكرة جديدة لنظام جوبيتر

الآن بعد أن أصبحت مفكرتك جاهزة، حان الوقت لكتابة وتشغيل برنامجك الأول في مفكرة جوبيتر.

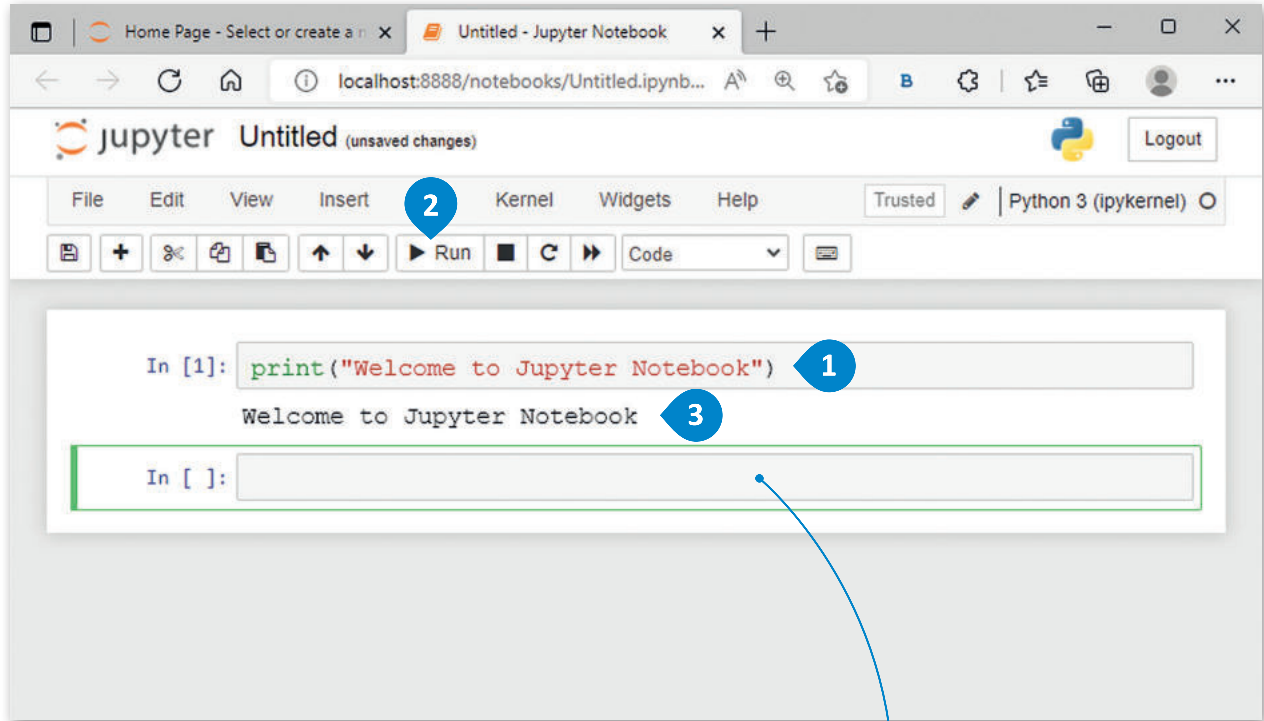
**لإنشاء برنامج في مفكرة جوبيتر:**

< اكتب **الأوامر** داخل خلية المقطع البرمجي. 1

< اضغط على Run (تشغيل). 2

< وسيتم عرض النتيجة تحت **الأوامر**. 3

**يمكنك تشغيل برنامجك بالضغط على Shift + Enter ↵.**




الشكل 3.8: إنشاء برنامج في مفكرة جوبيتر

**عند تشغيل برنامجك، ستتم إضافة خلية مقطع برمجي جديدة تلقائيًا.**

**يمكنك الحصول على العديد من الخلايا المختلفة التي تحتاجها في نفس المفكرة حيث تحتوي كل خلية على مقطعها البرمجي الخاص.**

**معلومة**

اسم مشروع جوبيتر هو إشارة إلى لغات البرمجة الأساسية الثلاث التي يدعمها جوبيتر، وهي جوليا (Julia) و البايثون (Python) و R.



حان الوقت لحفظ المفكرة الخاصة بك.

**لحفظ المفكرة الخاصة بك:**

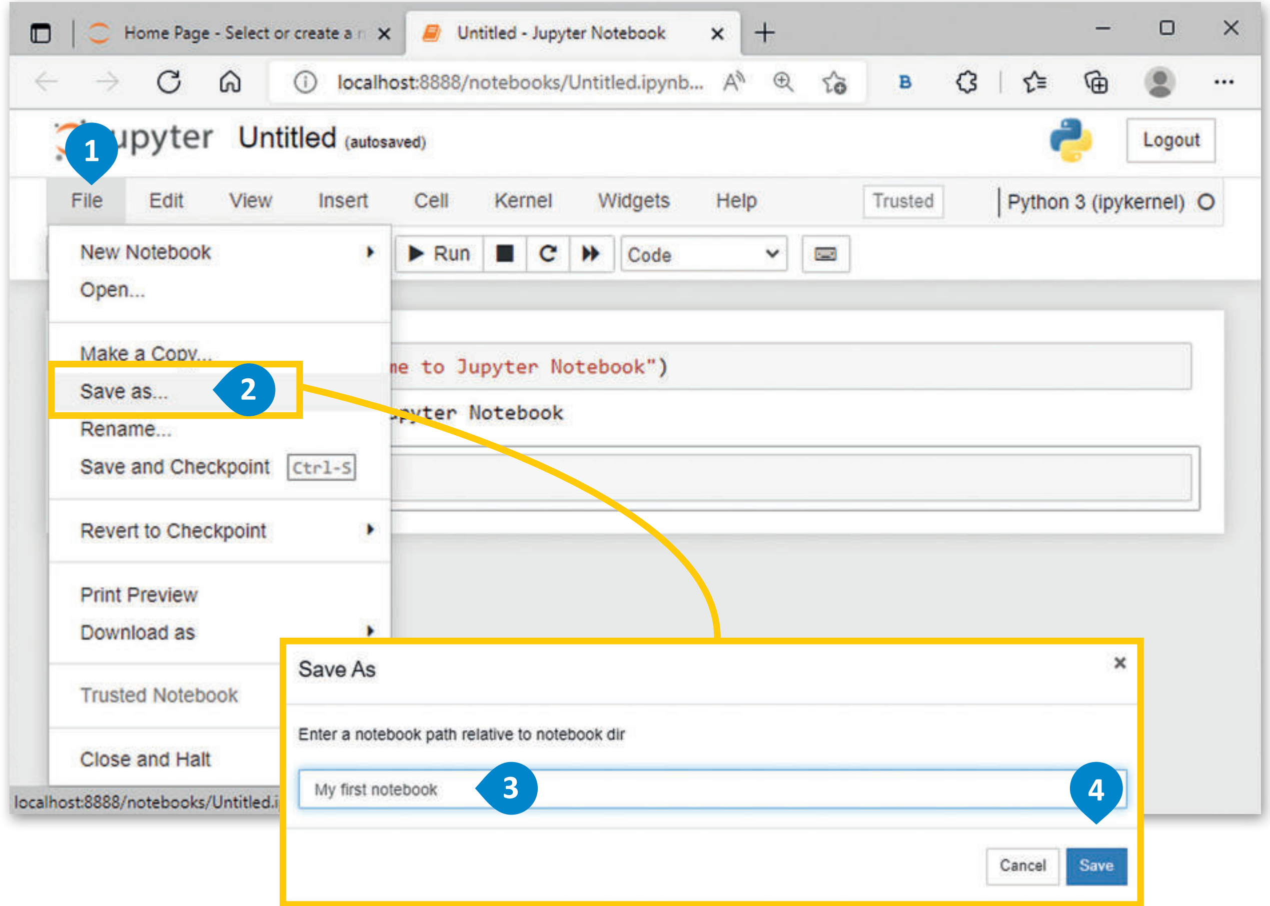
< اضغط على File (ملف). 1

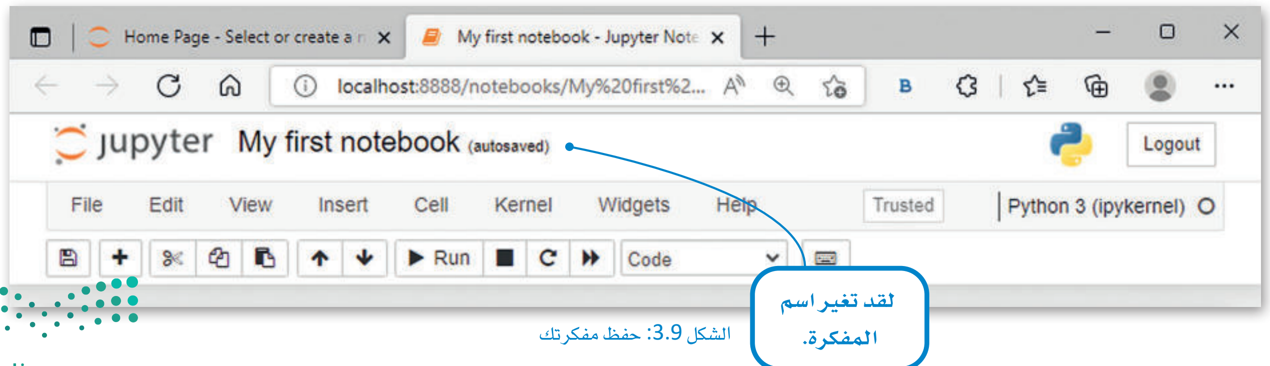
< اختر Save as (احفظ كـ). 2

< اكتب اسمًا لمفكرتك. 3

< اضغط على Save (حفظ). 4

**يتم حفظ المفكرة تلقائيًا أثناء عملك.**




الشكل 3.9: حفظ مفكرتك

# تمرينات

**1**

| حدد الجملة الصحيحة والجملة الخاطئة فيما يلي: | صحيحة | خاطئة |
|---|---|---|
| 1. يتم إجراء تحليل البيانات الوصفي لمعرفة سبب حدوث شيء ما. | ○ | ○ |
| 2. يوفر تحليل البيانات التشخيصي قيمة مضافة أكثر من تحليل البيانات التوجيهي. | ○ | ○ |
| 3. يستخدم تحليل البيانات التنبؤي التوجهات المكتشفة بالفعل للتنبؤ بالنتائج المستقبلية. | ○ | ○ |
| 4. تحليل البيانات التوجيهي هو أسهل أنواع تحليل البيانات. | ○ | ○ |
| 5. تحليل البيانات الاستكشافي هو دائمًا تمثيل بياني للبيانات. | ○ | ○ |
| 6. مع تحليل البيانات الاستكشافي، يمكنك اكتشاف الحالات الشاذة في مجموعة البيانات. | ○ | ○ |
| 7. يأخذ تحليل البيانات متعدد المتغيرات في الاعتبار أكثر من متغير مستقل واحد. | ○ | ○ |
| 8. تحتوي مكتبات البايثون على حزم من المقاطع البرمجية التي تبسّط العديد من مهام البرمجة. | ○ | ○ |
| 9. لا يمكن لمكتبة البايثون أن تحتوي على بيانات التكوين أو قوالب الرسائل. | ○ | ○ |
| 10. **مات بلوت ليب** (Matplotlib) هي إحدى مكتبات البايثون وتُستخدم لإنشاء الرسوم والمخططات البيانية. | ○ | ○ |



2 قارن بين تحليل البيانات التنبؤي والتوجيهي. ما الاختلافات؟ واذكر مثالاً على كل نوع من التحليل.

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

3 اذكر مثالين للمشاكل التي تتطلب تحليلاً أحادي المتغير ومثالين للمشاكل التي تتطلب تحليلاً متعدد المتغيرات، وهل يمكنك تحديد التعقيد المتزايد؟

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

4 اذكر إيجابيات وسلبيات استخدام مكتبات البايثون مقارنة بكتابة المقاطع البرمجية الخاصة بك، وما النهج الذي ستختاره؟

______________________________________________________________________

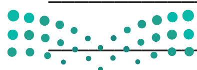
______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________



**5** أنت محلل بيانات في شركة تريد معرفة كيفية توزيع نفقاتها في مجالات مختلفة، فما نوع تحليل البيانات الذي ستطبقه؟ ولماذا؟

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**6** ما الميزة الرئيسية لاستخدام مفكرة جوبيتر؟

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**7** أنشئ مفكرة جديدة في جوبيتر:

- اطبع الرسالة "هذه مفكرتي الأولى".
- احفظ مفكرتك باسم من اختيارك.

______________________________________________

______________________________________________

______________________________________________

______________________________________________




رابط الدرس الرقمي
www.ien.edu.sa

الدرس الثاني

# مكتبات البايثون لتحليل البيانات

في الدرس السابق، تم التطرق إلى أن البايثون يستخدم المكتبات من أجل التعامل مع البيانات. في هذا الدرس ستتعلم طريقة توظيف مكتبات البايثون في مفكرة جوبيتر الخاصة بك.

## مكتبة نمباي NumPy Library

يرمز اسم مكتبة **نمباي** (NumPy) إلى **البايثون العددي** (Numerical Python)، وهي مكتبة قياسية للعمل مع البيانات العددية في البايثون، يمكن استخدامها لإجراء مجموعة متنوعة من العمليات الرياضية على المصفوفات.

**الجدول 3.3: وظائف مكتبة نيمباي**

| وظيفة | المعنى |
|---|---|
| add(arr1, arr2,...) | إضافة المصفوفات. |
| multiply(arr1,arr2,...) | ضرب المصفوفات. |
| absolute(arr) | تُرجع القيمة المطلقة لكل عنصر في المصفوفة المدخلة. |
| maximum(arr1,arr2,...) | تُرجع القيمة القصوى في المصفوفة المدخلة. |

**وظيفة (Method):**
هي دالة مرتبطة **بكائن** (Object) ويتم تعريفها داخل **الفئة** (Class). على سبيل المثال: np.add(arr1, arr2).

ابدأ بإنشاء قائمة بسيطة في مفكرة جوبيتر الخاصة بك. هذه قائمتك:

```
myList = [-3,-2,-1,0,1,2,3,4,5,5,5,6,7,8]

print(type(myList))
print(myList)

<class 'list'>
[-3, -2, -1, 0, 1, 2, 3, 4, 5, 5, 5, 6, 7, 8]
```

الشكل 3.10: وضع قائمة في مفكرة جوبيتر

**مصفوفة (Array):**
هي نوع من البيانات يمكنه الاحتفاظ بعدد ثابت من القيم التي لها نفس نوع البيانات.

استخدم مكتبة نمباي، وفي هذا المقطع البرمجي ستستخدم وظيفة **القيمة المطلقة (absolute())** لطباعة القيم المطلقة للقائمة.

```
import numpy as np

a = np.absolute(myList)
print(a)

[3 2 1 0 1 2 3 4 5 5 5 6 7 8]
```

عند استخدام مكتبة، يمكنك أن تعطيها اسمًا لاستخدام وظائفها في مقطعك البرمجي.

عند استخدام وظيفة من المكتبة، اكتب اسم المكتبة ثم نقطة ثم اسم الدالة.

الشكل 3.11: استخدام مكتبة نمباي

## مكتبة باندا س Pandas Library

تأخذ مكتبة باندا س البيانات وتنشئ كائن البايثون، وهناك نوعان رئيسيان من الكائنات:

> المتسلسلة (Series): عبارة عن مصفوفة أحادية البعد قادرة على حمل أي نوع من البيانات (الأعداد الصحيحة (Integers)، والسلاسل النصية (Strings)، والأرقام العشرية (Floats)، وكائنات البايثون وغيرها).

> إطار البيانات (DataFrame): هو هيكل بيانات ثنائي الأبعاد يبدو مشابهًا جدًا لجدول في ورقة عمل إكسل.

لكل كائن أساليبه وسماته الخاصة. يمكنك إنشاء متسلسلة أو إطار بيانات من الصفر (من القوائم والقواميس وما إلى ذلك) كما يمكن استيراد البيانات من مصادر البيانات، مثل إكسل و CSV، و SQL، و JSON، والمزيد.

**الجدول 3.4: الاختلافات بين مكتبات باندا س ونيمباي**

| | باندا س | نيمباي |
|---|---|---|
| **أنواع البيانات** | يعمل مع البيانات المجدولة. | يعمل مع البيانات العددية. |
| **أنواع الكائنات** | متسلسلة (Series)، إطار البيانات (DataFrame). | مصفوفات. |
| **الأداء** | يتعامل مع مئات الآلاف من البيانات. | يعالج خمسين ألف صف أو أقل. |
| **استخدام الذاكرة** | يستهلك المزيد من الذاكرة. | يستهلك ذاكرة أقل. |
| **الاستخدام** | تحليل البيانات وتصويرها. | إجراء الحسابات. |

## كائن المتسلسلة Series Object

الآن، ستقوم بتحويل هذه القائمة إلى كائن المتسلسلة. للقيام بذلك، عليك تضمين مكتبة باندا س في مفكرتك. ولاستخدام مكتبة في البايثون، يمكنك إضافة كلمة **استيراد** (Import) واسم المكتبة في بداية مقطعك البرمجي.

```
import pandas as pd
```

```
s = pd.Series(myList, name='Numbers')

print(s)
```

```
0     -3
1     -2
2     -1
3      0
4      1
5      2
6      3
7      4
8      5
9      5
10     5
11     6
12     7
13     8
Name: Numbers, dtype: int64
```

الشكل 3.12: تحويل قائمة إلى كائن المتسلسلة

**في مفكرة جوبيتر، عليك استيراد المكتبة مرة واحدة فقط ثم يمكنك استخدامها في المفكرة بأكملها.**



## سمات كائن المتسلسلة Attributes of Series Object

في الجدول 3.5 يتم تقديم بعض السمات الأكثر شيوعًا التي يمكنك استخدامها لكائن المتسلسلة.

**الجدول 3.5: سمات كائن المتسلسلة**

| السمة | المعنى |
|---|---|
| name | تُرجع اسم المتسلسلة. |
| size | تُرجع حجم المتسلسلة. |
| is_unique | تُرجع صواب (True) إذا كانت قيم كائن المتسلسلة فريدة، وإلا فإنها تُرجع **خطأ** (False). |
| hasnans | تُرجع صواب (True) إذا كان كائن المتسلسلة المعطى لديه قيم مفقودة، وإلا فإنها تُرجع **خطأ** (False). |

**السمة (Attribute):**

قيمة مرتبطة بالكائن الذي يشار إليه بالاسم باستخدام تعبيرات منقطة. على سبيل المثال، إذا كان الكائن **طالب** (student) وكانت السمة **درجة** (grade)، فسيتم الإشارة إليها student.grade.

**في الحوسبة، NaN ترمز إلى ليس رقمًا** (Not a Number).

طبّق بعض هذه السمات في كائن المتسلسلة.

```
# What is the name of the Series?
print("The name of the series is:", s.name)
```

```
The name of the series is: Numbers
```

```
# Print Series size
print("Size of the series is:", s.size)
```

```
Size of the series is: 14
```

```
print("Are the elements of this series unique?", s.is_unique)
```

```
Are the elements of this series unique? False
```

```
# Check if there are empty rows in the Series (nan = Not A Number)
print("Are there empty values in the series?", s.hasnans)
```

```
Are there empty values in the series? False
```

الشكل 3.13: السمات في كائن المتسلسلة

## كائن إطار البيانات DataFrame Object

الأداة التحليلية الأكثر شيوعًا واستخدامًا هي إكسل. يمكنك العمل مع ملفات إكسل في مفكرة جوبيتر باستخدام مكتبة باندا س. لفتح ملف إكسل في مفكرة جوبيتر، تحتاج إلى أن تكون هذه الملفات (ملف الإكسل والمفكرة) في نفس المجلد.

### مكتبة نظام التشغيل OS Library

للتحقق من ملف العمل الخاص بك، يمكنك استخدام مكتبة **نظام التشغيل** (OS)، حيث أنها توفر في بايثون وظائف لإنشاء وإزالة دليل (مجلد)، وجلب محتوياته، وتغيير أو تحديد المجلد الحالي، إلى آخره.

```
import os
os.getcwd()
```

```
'C:\\Users\\Documents\\Jupyter examples'
```

getcwd يرمز إلى
احصل على مجلد العمل الحالي
(get current working directory).

الشكل 3.14: مكتبة نظام التشغيل

هذا هو ملف الإكسل الخاص بك.

| | المنطقة الإدارية (A) | المرحلة (B) | نوع المدرسة (C) | مجموع الطلبة (D) |
|---|---|---|---|---|
| 2 | الرياض | التعليم المستمر | تعليم الكبار | 826 |
| 3 | الرياض | المرحلة الإبتدائية | تحفيظ | 1.040 |
| 4 | الرياض | المرحلة الإبتدائية | تربية خاصة | 190 |
| 5 | الرياض | المرحلة الإبتدائية | نهاري | 34.668 |
| 6 | الرياض | المرحلة الثانوية | تحفيظ | 285 |
| 7 | الرياض | المرحلة الثانوية | تربية خاصة | 71 |
| 8 | الرياض | المرحلة الثانوية | تعليم الكبار | 183 |
| 9 | الرياض | المرحلة الثانوية | نهاري | 16.018 |
| 10 | الرياض | المرحلة المتوسطة | تحفيظ | 548 |
| 11 | الرياض | المرحلة المتوسطة | تربية خاصة | 63 |
| 12 | الرياض | المرحلة المتوسطة | تعليم الكبار | 119 |
| 13 | الرياض | المرحلة المتوسطة | نهاري | 16.544 |
| 14 | الرياض | رياض الأطفال | تربية خاصة | 34 |
| 15 | الرياض | رياض الأطفال | نهاري | 3.836 |
| 16 | الرياض | المرحلة الإبتدائية | تحفيظ | 149 |
| 17 | الرياض | المرحلة الإبتدائية | نهاري | 6.057 |
| 18 | الرياض | المرحلة الثانوية | تعليم الكبار | 56 |
| 19 | الرياض | المرحلة الثانوية | نهاري | 2.216 |
| 20 | الرياض | المرحلة المتوسطة | تحفيظ | 10 |
| 21 | الرياض | المرحلة المتوسطة | نهاري | 1.481 |
| 22 | الرياض | رياض الأطفال | نهاري | 1.030 |
| 23 | الرياض | المرحلة الإبتدائية | نهاري | 380 |
| 24 | الرياض | المرحلة المتوسطة | نهاري | 40 |
| 25 | الرياض | رياض الأطفال | نهاري | 74 |

الطلبة والمعلمين والاداريين | المدارس | المدارس (مراحل) | فصول

الشكل 3.15: saudischools.xlsx ملف إكسل

**إن مجموعة البيانات التي ستستخدمها في هذا الدرس مُقدّمة من قبل وزارة التعليم من خلال المنصة السعودية للبيانات المفتوحة** https://data.gov.sa. **يمكنك استخدام بيانات ملف إكسل لغرض هذا الدرس باتباع سياسات البيانات المفتوحة** https://data.gov.sa/ar/policies.

الآن، ستقوم بتحويل ملف الإكسل التالي إلى إطار البيانات لمعالجة بياناته.

```
data = pd.read_excel('saudischools.xlsx')
```

```
data
```

| | المنطقة الإدارية | المرحلة | نوع المدرسة | مجموع الطلبة | مجموع المعلمين | مجموع الإداريين |
|---|---|---|---|---|---|---|
| 0 | الرياض | التعليم المستمر | تعليم الكبار | 826 | 0 | 0 |
| 1 | الرياض | المرحلة الإبتدائية | تحفيظ | 1040 | 72 | 41 |
| 2 | الرياض | المرحلة الإبتدائية | تربية خاصة | 190 | 123 | 10 |
| 3 | الرياض | المرحلة الإبتدائية | نهاري | 34668 | 1843 | 1152 |
| 4 | الرياض | المرحلة الثانوية | تحفيظ | 285 | 41 | 0 |
| ... | ... | ... | ... | ... | ... | ... |
| 5592 | نجران | المرحلة الإبتدائية | نهاري | 215 | 18 | 0 |
| 5593 | نجران | رياض الأطفال | نهاري | 154 | 40 | 0 |
| 5594 | نجران | المرحلة الإبتدائية | نهاري | 86 | 12 | 0 |
| 5595 | نجران | المرحلة المتوسطة | نهاري | 15 | 3 | 0 |
| 5596 | نجران | رياض الأطفال | نهاري | 19 | 2 | 0 |

5597 rows × 6 columns

الشكل 3.16: إنشاء إطار بيانات

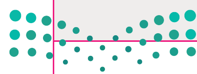
**إذا كان ملف الإكسل يحتوي على أوراق عمل متعددة، يمكنك قراءة ورقة عمل محددة. تستخدم وظيفة بانداس (read_excel) متغيرسمى (sheet_name) تخبر باندباس بورقة العمل التي يجب قراءتها من البيانات المُتاحة في ملف الإكسل. إذا لم تحدد الورقة، فإنها ستقرأ ورقة العمل الأولى.**

## سمات كائن إطار البيانات Attributes of a DataFrame Object

في الجدول التالي، يتم تقديم بعض السمات الأكثر شيوعًا، والتي يمكنك توظيفها في الحصول على معلومات حول إطار البيانات.

**الجدول 3.6: سمات كائن إطار البيانات**

| السمة | المعنى |
|---|---|
| shape | تُرجع أبعاد إطار البيانات. |
| size | تُرجع العدد الإجمالي للعناصر في إطار بيانات (n x m) |
| dtypes | تُرجع نوع القيمة لكل عمود. |
| columns | تُرجع أسماء أعمدة إطار البيانات. |
| axes | تُرجع عدد الصفوف وأسماء الأعمدة. |



```
# Printing the table dimensions
data.shape
```

```
(5597, 6)
```

```
# Return the total number of elements in the array (n x m)
data.size
```

```
33582
```

يمكنك إضافة تعليقات إلى مقطعك البرمجي باستخدام (#) في بداية العبارة.

```
# Return the type of the value of each column
data.dtypes
```

```
المنطقة الإدارية     object
المرحلة              object
نوع المدرسة          object
مجموع الطلبة         int64
مجموع المعلمين       int64
مجموع الإداريين      int64
dtype: object
```

```
# Return the number of rows and the name of the column
data.axes
```

```
[RangeIndex(start=0, stop=5597, step=1),
 Index(['المنطقة الإدارية', 'المرحلة', 'نوع المدرسة', 'مجموع الطلبة',
        'مجموع المعلمين', 'مجموع الإداريين'],
       dtype='object')]
```

الشكل 3.17: استخدام الخصائص على كائن إطار البيانات

في مكتبة بانداس، عادة ما يكون نوع بيانات الكائن سلسلة نصية string.data.

**الجدول 3.7: أنواع بيانات باندا س**

| نوع بيانات باندا س | نوع بيانات البايثون |
|---|---|
| object | str or mixed |
| int64 | int |
| float64 | float |
| bool | bool |
| datetime64 | NA |
| timedelta[ns] | NA |
| category | NA |



## الفهرسة Indexing

الفهرس (Index) هو قائمة بالأعداد الصحيحة أو التسميات التي تستخدمها لتحديد الصفوف أو الأعمدة بشكل فريد. في باندا س، تتضمن الفهرسة بشكل أساسي اختيار صفوف وأعمدة محددة من البيانات من إطار البيانات، حيث يمكن اختيار جميع الصفوف وبعض الأعمدة، أو اختيار بعض الصفوف وجميع الأعمدة، أو بعض من كل صف وعمود. اختيار **المجموعة الفرعية** (Subset Selection) هو مصطلح آخر للفهرسة.

لتشاهد بعض الأمثلة على الوظائف التي يمكنك استخدامها للفهرسة.

**الجدول 3.8: وظائف الفهرسة**

| الوظيفة | المعنى |
|---|---|
| head() | تُرجع العناصر الأولى من الكائن. |
| tail() | تُرجع العناصر الأخيرة من الكائن. |
| value_counts() | تُرجع القيم الفريدة للكائن. |
| idxmax() | تُرجع قيمة فهرس العنصر الأقصى. |
| idxmin() | تُرجع قيمة فهرس العنصر الأدنى. |

### استخدام الفهرسة في كائن المتسلسلة Using Indexing in a Series Object

طبّق وظائف الفهرسة هذه على كائن المتسلسلة الذي قمت بإنشائه. اطبع كائن المتسلسلة أولًا، لتذكر محتوياته.

```
print(s)
```

```
0     -3
1     -2
2     -1
3      0
4      1
5      2
6      3
7      4
8      5
9      5
10     5
11     6
12     7
13     8
Name: Numbers, dtype: int64
```

كائن المتسلسلة.

```
x=4
print("the value of the index",x, "is",s[x])
```

```
the value of the index 4 is 1
```

```
# Return the first 2 rows of the series
s.head(2)
```

```
0   -3
1   -2
Name: Numbers, dtype: int64
```

```
# Return the last rows of the series
s.tail()
```

```
9     5
10    5
11    6
12    7
13    8
Name: Numbers, dtype: int64
```

```
# Return a count of the unique values of the series
s.value_counts()
```

```
 5    3
-3    1
-2    1
-1    1
 0    1
 1    1
 2    1
 3    1
 4    1
 6    1
 7    1
 8    1
Name: Numbers, dtype: int64
```



الشكل 3.18: استخدام الفهرسة في كائن المتسلسلة



## استخدام الفهرسة في كائن إطار البيانات Using Indexing in DataFrame Object

```
# Printing the first 10 rows of the table
data.head(10)
```

| | المنطقة الإدارية | المرحلة | نوع المدرسة | مجموع الطلبة | مجموع المعلمين | مجموع الإداريين |
|---|---|---|---|---|---|---|
| **0** | الرياض | التعليم المستمر | تعليم الكبار | 826 | 0 | 0 |
| **1** | الرياض | المرحلة الإبتدائية | تحفيظ | 1040 | 72 | 41 |
| **2** | الرياض | المرحلة الإبتدائية | تربية خاصة | 190 | 123 | 10 |
| **3** | الرياض | المرحلة الإبتدائية | نهاري | 34668 | 1843 | 1152 |
| **4** | الرياض | المرحلة الثانوية | تحفيظ | 285 | 41 | 0 |
| **5** | الرياض | المرحلة الثانوية | تربية خاصة | 71 | 19 | 0 |
| **6** | الرياض | المرحلة الثانوية | تعليم الكبار | 183 | 0 | 0 |
| **7** | الرياض | المرحلة الثانوية | نهاري | 16018 | 980 | 462 |
| **8** | الرياض | المرحلة المتوسطة | تحفيظ | 548 | 42 | 34 |
| **9** | الرياض | المرحلة المتوسطة | تربية خاصة | 63 | 28 | 2 |

```
data.tail()
```

| | المنطقة الإدارية | المرحلة | نوع المدرسة | مجموع الطلبة | مجموع المعلمين | مجموع الإداريين |
|---|---|---|---|---|---|---|
| **5592** | نجران | المرحلة الإبتدائية | نهاري | 215 | 18 | 0 |
| **5593** | نجران | رياض الأطفال | نهاري | 154 | 40 | 0 |
| **5594** | نجران | المرحلة الإبتدائية | نهاري | 86 | 12 | 0 |
| **5595** | نجران | المرحلة المتوسطة | نهاري | 15 | 3 | 0 |
| **5596** | نجران | رياض الأطفال | نهاري | 19 | 2 | 0 |



```
# Acessing the dataframe attribute 'columns' to print the names of
# the table's columns
for col in data.columns:
    print(col)
```

```
المنطقة الإدارية
المرحلة
نوع المدرسة
مجموع الطلبة
مجموع المعلمين
مجموع الإداريين
```

طباعة أسماء أعمدة إطار البيانات.

```
data.describe()
```

تستخدم وظيفة describe() لعرض بعض التفاصيل الإحصائية الأساسية.

| | مجموع الطلبة | مجموع المعلمين | مجموع الإداريين |
|---|---|---|---|
| count | 5597.000000 | 5597.000000 | 5597.000000 |
| mean | 1110.124352 | 89.510988 | 19.459175 |
| std | 2950.990275 | 192.359535 | 66.800341 |
| min | 0.000000 | 0.000000 | 0.000000 |
| 25% | 31.000000 | 4.000000 | 0.000000 |
| 50% | 136.000000 | 17.000000 | 1.000000 |
| 75% | 808.000000 | 82.000000 | 10.000000 |
| max | 36416.000000 | 2090.000000 | 1152.000000 |

الشكل 3.19: استخدام الفهرسة في كائن إطار البيانات



## تصفية البيانات أو اختيار مجموعة بيانات جزئية
## Filtering Data or Subset Selection

في بعض الأحيان لا تحتاج إلى مجموعة البيانات بأكملها. تحتاج إلى عزل بعض البيانات المحددة. للقيام بذلك، تحتاج إلى إضافة بعض المرشحات. هناك العديد من الأساليب لاختيار مجموعة جزئية من إطار البيانات أو متسلسة. الأسلوب الأسهل هو استخدام الفهرسة المنطقية (Boolean Indexing)، ولكن الأسلوب الأكثر قوة هو باستخدام طرق loc و iloc. أولاً ستتعلم الفهرسة المنطقية ثم أسلوب loc و iloc.

**تصفية البيانات:**

تصفية البيانات هو عملية اختيار جزء أصغر من مجموعة البيانات الخاصة بك واستخدام تلك المجموعة الجزئية للعرض أو التحليل.

### الفهرسة المنطقية Boolean Indexing

هي نوع من الفهرسة التي تستخدم القيم الفعلية لمجموعة البيانات، وفيها تحتاج إلى استخدام **المُعاملات المنطقية** (Boolean Operator)، وتُكتب المُعامِلات المنطقية بشكل مختلف في مفكرة جوبيتر عن بايثون.

**الجدول 3.9: المُعامِلات المنطقية في مفكرة جوبيتر**

| جوبيتر | بايثون |
|---|---|
| & | and |
| \| | or |
| ~ | not |

لتشاهد بعض الأمثلة على كائن المتسلسلة.

```
# Return the elements of the series that satisfy the expression s>0
s[s > 0]

4     1
5     2
6     3
7     4
8     5
9     5
10    5
11    6
12    7
13    8
Name: Numbers, dtype: int64

s[(s < -1) | (s > 6)]

0    -3
1    -2
12    7
13    8
Name: Numbers, dtype: int64

# Printing not(s<0)  =>  (s>=0)
s[~(s < 0)]

3     0
4     1
5     2
6     3
7     4
8     5
9     5
10    5
11    6
12    7
13    8
Name: Numbers, dtype: int64
```

الشكل 3.20: تصفية البيانات في الكائنات المتسلسلة



## الفهرسة مع أسلوب Loc و Iloc Indexing with Loc and Iloc Methods

تُعد طريقتي iloc و loc ضمن الطرق الأكثر شيوعًا للفهرسة في مكتبة بانداس.

< **loc**: يختار الصفوف والأعمدة مع مسميات محددة (أسماء الأعمدة).

< **iloc**: يختار الصفوف والأعمدة في مواضع الأعداد الصحيحة المحددة (أرقام الصفوف والأعمدة).

وإليك أدناه بعض الأمثلة على استخدام كائن إطار البيانات بأسلوب loc().

في هذا المثال، ستستخدم طريقة loc() لطباعة الصفوف الخمسة الأولى من عمودين محددين.

```
# Choosing the first 5 rows of the columns 'المنطقة الإدارية' and 'المرحلة'
data.loc[:4,['المرحلة','المنطقة الإدارية']]
```

| | المنطقة الإدارية | المرحلة |
|---|---|---|
| 0 | الرياض | التعليم المستمر |
| 1 | الرياض | المرحلة الإبتدائية |
| 2 | الرياض | المرحلة الإبتدائية |
| 3 | الرياض | المرحلة الإبتدائية |
| 4 | الرياض | المرحلة الثانوية |

الشكل 3.21: طباعة الصفوف الخمسة الأولى من عمودين محددين

في هذا المثال، ستقوم بطباعة صفوف إطار البيانات التي لها قيمة محددة في عمود معين.

```
# Print the rows of the dataframe that have a specific value in a specific column
data.loc[data['المنطقة الإدارية'].isin(['الرياض','الباحة'])]
```

| | المنطقة الإدارية | المرحلة | نوع المدرسة | مجموع الطلبة | مجموع المعلمين | مجموع الإداريين |
|---|---|---|---|---|---|---|
| 0 | الرياض | التعليم المستمر | تعليم الكبار | 826 | 0 | 0 |
| 1 | الرياض | المرحلة الإبتدائية | تحفيظ | 1040 | 72 | 41 |
| 2 | الرياض | المرحلة الإبتدائية | تربية خاصة | 190 | 123 | 10 |
| 3 | الرياض | المرحلة الإبتدائية | نهاري | 34668 | 1843 | 1152 |
| 4 | الرياض | المرحلة الثانوية | تحفيظ | 285 | 41 | 0 |
| ... | ... | ... | ... | ... | ... | ... |
| 4807 | الباحة | المرحلة الثانوية | نهاري | 391 | 54 | 3 |
| 4808 | الباحة | المرحلة المتوسطة | تحفيظ | 17 | 10 | 0 |
| 4809 | الباحة | المرحلة المتوسطة | تربية خاصة | 1 | 0 | 0 |
| 4810 | الباحة | المرحلة المتوسطة | نهاري | 387 | 62 | 3 |
| 4811 | الباحة | رياض الأطفال | نهاري | 220 | 18 | 10 |

1318 rows × 6 columns

الشكل 3.22: طباعة صفوف إطار البيانات التي لها قيمة محددة في عمود محدد

في هذا المثال، ستنشئ إطار بيانات جديدًا يسمى studentsReg. وسيحتوي إطار البيانات هذا على عمودين: عمود واحد للمنطقة وآخر لعدد الطلبة.

```
# Create a dataframe called studentsReg with two columns Region and Number of Students
studentsReg = data.loc[:,['مجموع الطلبة','المنطقة الإدارية']]
studentsReg
```

| | المنطقة الإدارية | مجموع الطلبة |
|---|---|---|
| 0 | الرياض | 826 |
| 1 | الرياض | 1040 |
| 2 | الرياض | 190 |
| 3 | الرياض | 34668 |
| 4 | الرياض | 285 |
| ... | ... | ... |
| 5592 | نجران | 215 |
| 5593 | نجران | 154 |
| 5594 | نجران | 86 |
| 5595 | نجران | 15 |
| 5596 | نجران | 19 |

5597 rows × 2 columns

الشكل 3.23: إنشاء إطار بيانات جديد يسمى studentsReg

والآن، سوف تستخدم طريقة iloc() لتحديد جميع عناصر الصف الأول من إطار البيانات.

تذكر، الفهرسة في بايثون تبدأ من 0.

```
# Print all the elements from the [row] of the table
data.iloc[0]
```

```
المنطقة الإدارية          الرياض
المرحلة                 التعليم المستمر
نوع المدرسة              تعليم الكبار
مجموع الطلبة             826
مجموع المعلمين           0
مجموع الإداريين          0
Name: 0, dtype: object
```

الشكل 3.24: طباعة عناصر الصف الأول من إطار البيانات



وفي الأمثلة أدناه، ستطبع عناصر محددة من إطار البيانات.

```
# Print the element in the [row,col] position of the table
data.iloc[0,3]
```

```
826
```

```
# Print the elements [start:end, start:end], the end is not included
data.iloc[1:3, 0:3]
```

| | المنطقة الإدارية | المرحلة | نوع المدرسة |
|---|---|---|---|
| 1 | الرياض | المرحلة الإبتدائية | تحفيظ |
| 2 | الرياض | المرحلة الإبتدائية | تربية خاصة |

طباعة العناصر في الصفين الثاني والثالث، ولكن فقط من الأعمدة 0 و 1 و 2.

الشكل 3.25: طباعة عناصر محددة من إطار البيانات

وفي هذا المثال، سوف تستخدم حلقة **for** لطباعة الصفوف العشرة الأولى من العمود الأول من إطار بيانات studentsReg.

```
for i in range (10):
    print(studentsReg.iloc[i][1])
```

```
826
1040
190
34668
285
71
183
16018
548
63
```

الشكل 3.26: العناصر المطبوعة لإطار البيانات



## المجموعات والتجميع Grouping and Aggregating

تسمى عملية وضع عناصر مجموعة البيانات في مجموعات بناءً على بعض المعايير وتطبيق الوظائف على هذه المجموعات بالتجميع. في مكتبة باندا س؛ يتم تنفيذ هذا الإجراء باستخدام وظيفة (df.groupby()).

فعلى سبيل المثال، تخيل أن لديك مجموعة بيانات لأفضل هدّاف كرة السلة في كل العصور. إذا كنت ترغب في معرفة عدد اللاعبين في مجموعة البيانات هذه لفريق معين، فيمكنك تجميع هذه البيانات حسب عمود "الفريق" وتطبيق دالة المجموع (sum()) على البيانات المجمّعة.

**دالة التجميع:**

دالة تقوم بحسابات رياضية مع قيم صفوف متعددة والتي يتم تجميعها معًا، ونتيجة لذلك ترجع قيمة موجزة واحدة.

دوال التجميع الأكثر شيوعًا هي sum، count، max، min and mean.

### وظيفة Groupby
### Groupby Method

باستخدام وظيفة groupby() يمكنك تقسيم بياناتك إلى مجموعات مختلفة، ويمكن أن يساعدك هذا في إجراء حسابات لتحليل البيانات بشكل أفضل.

**الجدول 3.10: الدوال التجميعية**

| الدالة | المعنى |
|---|---|
| sum | تُرجع مجموع قائمة الأرقام. |
| max | تُرجع العدد الأقصى لقائمة الأرقام. |
| min | تُرجع العدد الأدنى لقائمة الأرقام. |
| mean | تُرجع متوسط قائمة الأرقام. |

لتشاهد بعض الأمثلة لوظيفة df.groupby() في إطار البيانات الجديد الذي أنشأته.

في هذا المثال، ستقوم بتجميع الطلبة وفقًا لمنطقتهم وتحسب مجموع الطلبة في كل منطقة.

```
# Create a dataframe with two rows, one being Regions and
# the other one the total number of students of each region
studentsReg.groupby('المنطقة الإدارية').sum()
```

| المنطقة الإدارية | مجموع الطلبة |
|---|---|
| الباحة | 84464 |
| الجوف | 151527 |
| الحدود الشمالية | 88482 |
| الرياض | 1473236 |
| الشرقية | 909884 |
| القصيم | 292996 |
| المدينة المنورة | 443254 |
| تبوك | 205673 |
| جازان | 354608 |
| حائل | 162609 |
| عسير | 480060 |
| مكة المكرمة | 1421176 |
| نجران | 145397 |

دالة sum هي إحدى دوال التجميع التي تدعمها مكتبة باندا س.

الشكل 3.27: استخدام وظيفة df.groupby() لتجميع إطار البيانات بعمود واحد

في هذا المثال، ستقوم بتجميع الطلبة وفقًا لمعيارين، منطقتهم ومرحلتهم الدراسية، وتحسب مجموع الطلبة في كل منطقة.

```
# Group by two columns, first group the table by Region and
# then by Level and for each group calculate the total NumOfStud
data.groupby(['المرحلة','المنطقة الإدارية'],as_index=False)['مجموع الطلبة'].sum()
```

| | المنطقة الإدارية | المرحلة | مجموع الطلبة |
|---|---|---|---|
| 0 | الباحة | التعليم المستمر | 553 |
| 1 | الباحة | المرحلة الإبتدائية | 40495 |
| 2 | الباحة | المرحلة الثانوية | 18386 |
| 3 | الباحة | المرحلة المتوسطة | 19158 |
| 4 | الباحة | رياض الأطفال | 5872 |
| ... | ... | ... | ... |
| 60 | نجران | التعليم المستمر | 1941 |
| 61 | نجران | المرحلة الإبتدائية | 75462 |
| 62 | نجران | المرحلة الثانوية | 29097 |
| 63 | نجران | المرحلة المتوسطة | 31688 |
| 64 | نجران | رياض الأطفال | 7209 |

65 rows × 3 columns

مجموعة المنطقة الأولى

مجموعة المنطقة الأخيرة

ضبط as_index=false ليكون الفهرس هو رقم الصف بدلاً من أسماء الأعمدة.

الشكل 3.28: استخدام وظيفة df.groupby() لتجميع إطار البيانات في أعمدة متعددة

وفي هذا المثال، ستقوم بإنشاء إطار بيانات جديد للمنطقة وعدد الطلبة وعدد المعلمين. ثم تجمع حسب المنطقة وتحسب مجموع الطلبة ومجموع المعلمين في كل منطقة.

```
# Create a table with one column for Regions, one column for total #Students and one for total #Teachers
studentsTeachers = data.loc[:, ['المنطقة الإدارية','مجموع الطلبة','مجموع المعلمين']]
studentsTeachers = studentsTeachers.groupby('المنطقة الإدارية')[['مجموع الطلبة','مجموع المعلمين']].sum()
studentsTeachers
```

| المنطقة الإدارية | مجموع الطلبة | مجموع المعلمين |
|---|---|---|
| الباحة | 84464 | 11092 |
| الجوف | 151527 | 12503 |
| الحدود الشمالية | 88482 | 7156 |
| الرياض | 1473236 | 114635 |
| الشرقية | 909884 | 62982 |
| القصيم | 292996 | 29971 |
| المدينة المنورة | 443254 | 37617 |
| تبوك | 205673 | 16048 |
| جازان | 354608 | 26487 |
| حائل | 162609 | 17289 |
| عسير | 480060 | 46865 |
| مكة المكرمة | 1421176 | 108309 |
| نجران | 145397 | 10039 |

الشكل 3.29: استخدام وظيفة df.groupby() وتحديد العمود المصنف كمؤشر للبيانات الجديدة



## تنظيف البيانات Data Cleaning

من المهم جدًا أن تكون البيانات التي ستحللها صحيحة ، قبل البدء بتحليلها، وهذا يعني أنه يجب إزالة البيانات المكررة أو المشوّشة أو غير الدقيقة من مجموعة البيانات الخاصة بك، وإذا بقيت هذه البيانات كما هي، فلن تكون نتائج تحليلها صحيحة.

**تنظيف البيانات:**

تنظيف البيانات هو عملية إصلاح أو إزالة البيانات غير الصحيحة أو المشوّشة أو المنسقة بشكل غير صحيح أو المكررة أو غير المكتملة من مجموعة البيانات.

إصلاح البيانات

إصلاح الخلايا الفارغة

إزالة البيانات المكررة

الشكل 3.30: عملية تنظيف البيانات

**الجدول 3.11: وظائف تنظيف البيانات**

| الوظيفة | المعنى |
|---|---|
| duplicated() | تُرجع قيمة منطقية لكل صف يحتوي على بيانات مكررة. |
| value_counts() | تُرجع القيم الفريدة في مجموعة البيانات. |
| isnull() | تُرجع قيمة منطقية لكل خلية فارغة من مجموعة البيانات. |
| dropna() | يحذف الصفوف الفارغة. |

## البيانات المكررة Duplicated Data

للتحقق مما إذا كانت مجموعة البيانات الخاصة بك تحتوي على بيانات مكررة، فيمكنك أن تستخدم الوظيفة df.duplicated() وتعطي هذه الوظيفة قيمة منطقية لكل صف حسب احتواءه على بيانات مكررة.

- صواب (True) للبيانات المكررة.
- خطأ (False) للبيانات غير المكررة.

سترى كيفية التعامل مع الصفوف المكررة في مجموعة البيانات.

```
dup = data.duplicated()
```

```
# To see how many duplicated rows there are in the table
dup.value_counts()
```

```
False    5426
True      171
dtype: int64
```

عدد النسخ المكررة

الشكل 3.31: استخدام وظيفة df.duplicated()

يوجد في مجموعة البيانات الخاصة بك 171 صفًا مكررًا.

لحذف هذه الصفوف تستخدم وظيفة ()drop_duplicates، حيث تحذف هذه الطريقة الصفوف المكررة.

بعد حذف الصفوف المكررة، عليك تحديث مجموعة البيانات الخاصة بك للتحقق من إزالة الصفوف المكررة.

```
# Now remove duplicated rows from the table
data = data.drop_duplicates()

dup = data.duplicated()

dup.value_counts()

False     5426
dtype: int64
```

لا يوجد صفوف مكررة.

الشكل 3.32: استخدام وظيفة ()drop_duplicates

## الخلايا الفارغة Empty Cells

للتحقق مما إذا كانت مجموعة البيانات الخاصة بك بها قيم مفقودة، يمكنك استخدام وظيفة ()data.isnull، حيث ترجع قيمة منطقية لكل خلية من مجموعة البيانات:

- صواب (True) للخلايا الفارغة
- خطأ (False) للخلايا الممتلئة

سترى كيف يمكنك عد الخلايا الفارغة في مجموعة البيانات.

في هذا المثال تحسب الخلايا الفارغة لكل عمود.

```
# get the number of empty cells per column
missing_values_count = data.isnull().sum()
missing_values_count

5    المنطقة الإدارية
6    المرحلة
5    نوع المدرسة
4    مجموع الطلبة
4    مجموع المعلمين
4    مجموع الإداريين
dtype: int64
```

عدد الخلايا الفارغة في كل عمود.

الشكل 3.33: عد الخلايا الفارغة لكل عمود

يمكنك رؤية عدد الخلايا الفارغة في كل عمود.

لحذف هذه الصفوف، تستخدم وظيفة ()dropna، وستقوم بحذف الصفوف الفارغة.

بعد حذف الصفوف الفارغة، عليك تحديث مجموعة البيانات الخاصة بك للتحقق من إزالة هذه الصفوف.



```
# Drop the missing values
data = data.dropna()
```

```
missing_values_count = data.isnull().sum()
missing_values_count
```

```
0      المنطقة الإدارية
0      المرحلة
0      نوع المدرسة
0      مجموع الطلبة
0      مجموع المعلمين
0      مجموع الإداريين
dtype: int64
```

لا يوجد خلايا فارغة

الشكل 3.34: حذف الصفوف الفارغة

## البيانات الخاطئة Wrong Data

في بعض الأحيان قد تحتوي مجموعة البيانات الخاصة بك على بيانات خاطئة. فعلى سبيل المثال، في مجموعة البيانات الخاصة بك لا يمكنك الحصول على أرقام سالبة في عدد عمود الطلبة، وللتحقق مما إذا كانت مجموعة البيانات الخاصة بك تحتوي على بيانات خاطئة، عليك كتابة مقطع برمجي مخصص على حسب مجموعة البيانات الخاصة بك.

في هذا المثال سوف تتحقق من الأرقام السالبة في أعمدة مجموعة البيانات.

يعتمد نوع البيانات التي يمكن اعتبارها خاطئة على مجموعة البيانات. عليك أن تقرر ماذا تفعل بهذه البيانات الخاطئة، فقد ترغب في حذفها أو استبدالها بقيم أخرى.

```
# Check if there are negative elements in the columns that have numbers
data[data['مجموع الطلبة'] <0].nunique()
```

```
0      المنطقة الإدارية
0      المرحلة
0      نوع المدرسة
3      مجموع الطلبة
3      مجموع المعلمين
3      مجموع الإداريين
dtype: int64
```

```
data[data['مجموع المعلمين'] <0].nunique()
```

```
0      المنطقة الإدارية
0      المرحلة
0      نوع المدرسة
3      مجموع الطلبة
3      مجموع المعلمين
3      مجموع الإداريين
dtype: int64
```

```
data[data['مجموع الإداريين'] <0].nunique()
```

```
0      المنطقة الإدارية
0      المرحلة
0      نوع المدرسة
3      مجموع الطلبة
3      مجموع المعلمين
3      مجموع الإداريين
dtype: int64
```

الشكل 3.35: التحقق من الأرقام السالبة

# تمرينات

**1** قارن بين كائن المتسلسلة و إطار البيانات.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**2** صف الفرق بين فهرسة البيانات وتصفيتها.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**3** اشرح أهمية تنظيف البيانات قبل البدء بتحليلها.

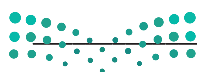
______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________



**4** استورد المكتبة العشوائية واستخدم وظيفة ()random.randrange لطباعة رقم عشوائي بين 1 و 100.

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

**5** افتح مفكرة جديدة في جوبيتر، واستورد ملف إكسل بالاسم "tourist-indicators.xlsx".

- حمّل ورقة العمل "I8" في إطار بيانات باندا س.
- اطبع شكل إطار البيانات.
- اطبع أنواع البيانات المخزنة في كل عمود من مجموعة البيانات.
- اطبع أسماء أعمدة مجموعة البيانات.



**6** استخدم مجموعة البيانات التي استوردتها في التمرين السابق وقم بالتالي:

- أنشئ كائن متسلسلة جديد يحتوي على عدد السياح الوافدين من دول مجلس التعاون الخليجي.
- أوجد الحد الأقصى والحد الأدنى لعدد السياح وفي أي صفوف من المتسلسلة تحدث هذه القيم؟
- تحقق من المتسلسلة لمعرفة القيم غير المناسبة والمفقودة، وإذا كان هناك أي منها، قم بإزالة تلك الصفوف.
- اطبع عدد السياح بالترتيب التنازلي لقيم أكبر من 500.

**7** افتح ورقة العمل "I8" من الملف "tourist-indicators.xlsx" واقرأها في إطار بيانات جديد. ثم:

- تَفقّد مجموعة البيانات بأكملها لمعرفة القيم المكررة والمفقودة.
- اطبع عدد القيم المفقودة وعدد الصفوف المكررة.
- قم بإزالة الصفوف المكررة والصفوف ذات القيم المفقودة.
- قم بتجميع إطار البيانات بناء على الشهر، وحدد الشهر الذي استقطب أكبر عدد من الزوار لكل منطقة.






الدرس الثالث

# تصوير البيانات

كما ذكر سابقًا، فإن تصوير البيانات هو التمثيل البياني للمعلومات والبيانات. إن تصوير البيانات يجعلها أيسر فهمًا وتحليلًا. باستخدام العناصر المرئية مثل المخططات والرسوم البيانية والخرائط، فإنك تجعل البيانات أكثر سهولة وفهمًا وقابلية للاستخدام.

في هذا الدرس، ستستخدم مفكرة جوبيتر لتصوير بياناتك. ويدعم جوبيتر تصوير البيانات بالاستعانة بمكتبات البايثون.

## أنواع تصوير البيانات
## Types of Data Visualization

أكثر أنواع تصوير البيانات شيوعًا هي:

- المخططات (الخطية، الشريطية، الدائرية)
- الرسوم البيانية
- المخطط النقطي
- المخطط المدرّج التكراري
- الجداول
- الخرائط

**يتم تمثيل البيانات بشكل مختلف باستخدام الأنواع المختلفة لتصوير البيانات.. يجب عليك اختيار نوع الرسم البياني حسب ما تريد تحقيقه من تقريرك.**

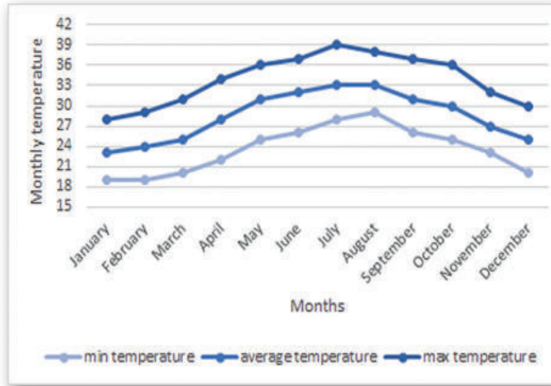


الشكل 3.36: الرسوم الأكثر شيوعا في تصوير البيانات

## المخططات Charts

### المخطط الخطّي Line Chart

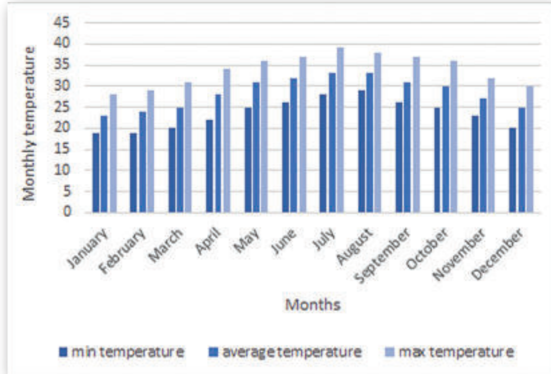
المخطط الخطّي هو تقنية تصوير بيانات، بحيث يتم رسم كل قيمة لمتغير مستقل على مدى فترة زمنية وتتصل هذه القيم بخطوط مستقيمة. عادة ما يكون المحور الأفقي متغيرًا مستمرًا مثل الوقت، والمحور الرأسي هو قيم المتغير المستقل. وتكمن بعض المزايا في بساطته في تمثيل تغيير المتغير بمرور الوقت والذي يمكن أن يساعد في اكتشاف التوجهات والأنماط. ويمكنك رسم خطوط متعددة على نفس الرسم البياني ومقارنة تقدم أكثر من متغير مستقل واحد في نفس الفترة الزمنية.




الشكل 3.37: مخطط خطّي يوضح المتوسط السنوي لدرجات الحرارة المنخفضة والمتوسطة المسجلة في أبْهَا

### المخطط الشريطي Bar Chart

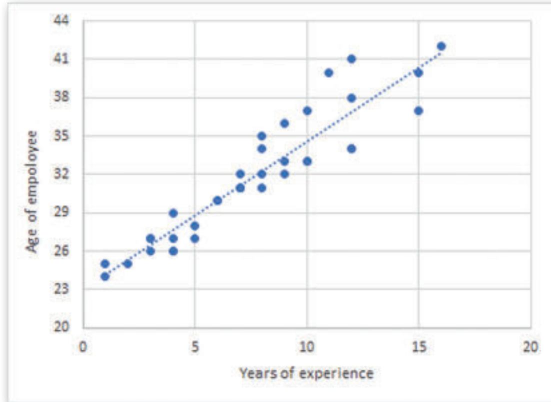
المخطط الشريطي يمثّل عناصر متغير فئوي على المحور الأفقي (س)، بينما توضح الأعمدة قيم تلك العناصر من خلال ارتفاعها نسبة إلى قيم المحور الرأسي (ص). يمكن أن تكون المخططات الشريطية عمودية أو أفقية، وعادة ما تسمى المخططات الشريطية العمودية مخططات الأعمدة. وهناك العديد من أنواع المخططات الشريطية مثل المخططات الشريطية المجمّعة، والمخططات الشريطية المكدّسة، والمخططات الشريطية مع أشرطة الخطأ، وغيرها المزيد.




الشكل 3.38: مخطط شريطي يوضح المتوسط السنوي لدرجات الحرارة المنخفضة والمتوسطة المسجلة في أبْهَا

### المخطط النقطي Scatter Plot

المخطط النقطي هو طريقة لتصوير البيانات باستخدام النقاط لتمثيل قيم المتغيرات المختلفة، وتكون هذه النقاط مبعثرة على الشكل، ومن هنا جاء الاسم. موقع هذه النقط على محوري (س) و (ص) يمثل قيمها، ويمكنك استخدام ألوان مختلفة لرسم النقاط، حيث يمثل كل لون متغير معين. وعندما تكون قيم المتغيرات التي تمت دراستها بيانات متقطعة، فإن المخطط النقطي يكون أكثر ملاءمة من المخطط الخطّي، حيث أنه أكثر قابلية للتطبيق لتمثيل المتغيرات ذات القيم المستمرة (الحقيقية). وهناك أنواع مختلفة من المخطط النقطي بناءً على الارتباط بين المتغيرات (ايجابي، سلبي، لاغي).

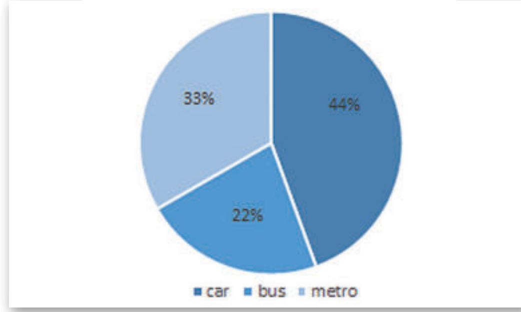



الشكل 3.39: مخطط نقطي يبين وجود ارتباط إيجابي بين سنوات الخبرة وعمر الموظف



## المخطط الدائري Pie Chart

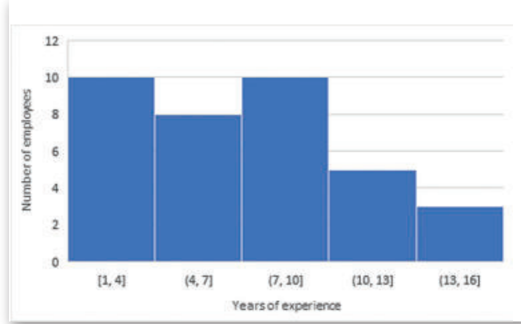
المخطط الدائري هو مخطط يشبه الفطائر، مقسّم إلى شرائح تمثل القيم النسبية لبعض المتغيرات في فئة معينة. تمثل كل شريحة من المخطط فئة مختلفة. هناك العديد من أنواع المخططات الدائرية، مثل **المخططات الدائرية المجوّفة** (Doughnut Charts) و**المخططات نصف المجوّفة** (Half-Doughnut Pie Charts) و**المخططات الدائرية متعددة الطبقات** (Multilayered Pie Charts).




الشكل 3.40:مخطط دائري يبين النسبة المئوية لوسيلة النقل المفضلة

## المخطط المدرّج التكراري Histogram

يعد المخطط المدرّج التكراري (الهيستوقرام) أحد أقدم تقنيات تصوير البيانات، حيث يشبه المخططات الشريطية ولكنه يختلف عنها في أنه يظهر تواتر البيانات العددية، بينما المخططات الشريطية تُعد طريقة لمقارنة فئات البيانات. وعندما تريد إنشاء مخطط المدرّج التكراري، فعليك بتجميع البيانات في نطاقات يتم رسمها بعد ذلك على شكل أعمدة متصلة ببعضها البعض، ويُظهر ارتفاع الأعمدة عدد البيانات الموجودة في كل نطاق.




الشكل 3.41: مخطط مدرّج تكراري يُبيّن توزيع سنوات الخبرة بين الموظفين

**البيانات الفئوية هي متغيرات متقطعة، ويمكن أن يكون لها عدد معين من القيم، فعلى سبيل المثال عدد الطلبة في كل منطقة من المملكة العربية السعودية. ويمكن أن يكون للبيانات المستمرة أي قيمة بين الحد الأدنى والقيمة القصوى، على سبيل المثال، الوقت أو درجة الحرارة.**

## مكتبة مات بلوت ليب Matplotlib Library

من أجل تصوير بياناتك، تحتاج إلى استيراد مكتبة جديدة، وهي التي تسمى مات بلوت ليب. وتحتوي هذه المكتبة على بعض الأساليب الجاهزة التي يمكنك استخدامها لجعل المخطط الخاص بك أكثر قابلية للفهم، ويمكنك الاطلاع على هذه الأساليب في الجدول 3.12. وباستخدام هذه المكتبة، يمكنك تقديم بياناتك في أي مخطط تريده. في هذا الدرس، ستستخدم هذه الأساليب لإنشاء مخططات بناءً على إطار البيانات الخاص بك.

**الجدول 3.12: طرق مكتبة مات بلوت ليب** (Matplotlib)

| الطريقة | المعنى |
|---|---|
| bar() | ينشئ مخطط شريطي. |
| pie() | ينشئ مخطط دائري. |
| set_title() | يحدد عنوان المخطط. |
| set_ylabel() | يحدد تسمية محور y. |
| set_xlabel() | يحدد تسمية محور x. |
| show() | ينشئ المخطط. |

لدعم النص العربي داخل المخططات التي أنشأتها مكتبة مات بلوت ليب، تحتاج إلى تحويل النص العربي إلى تنسيق يمكن عرضه بشكل صحيح. ستستخدم مكتبتي البايثون:

> arabic_reshaper

> bidi.algorithm

من خلال تشغيل المقطع البرمجي التالي في مفكرة جوبيتر الخاصة بك، يتم تنزيل هاتين المكتبتين وتثبيتهما تلقائيًا.

```
!pip install arabic-reshaper
!pip install python-bidi
```

الشكل 3.42: تنزيل وتثبيت المكتبات



## المخطط الشريطي Bar Chart

حان الوقت لإنشاء أول مخطط لك في مفكرة جوبيتر.

ابدأ باستيراد المكتبات التي ستستخدمها.

```
import matplotlib.pyplot as plt
import numpy as np
import arabic_reshaper
from bidi.algorithm import get_display
```

الشكل 3.43: المكتبات المستوردة

الخطوة التالية هي إنشاء مجموعة البيانات التي ستستخدمها.

من مجموعة البيانات التي استخدمتها في الدرس السابق، ابدأ بتجميع بياناتك حسب المنطقة باستخدام دالة mean() واحصل على متوسط عدد الطلبة والمعلمين والإداريين.

قم بفرز إطار البيانات هذا حسب عمود الطلبة.

```
groupsB = data.groupby(['المنطقة الإدارية'],as_index=False) [['مجموع الإداريين','مجموع المعلمين','مجموع الطلبة']].mean().round(0)
groupsB = groupsB.sort_values(by=['مجموع الطلبة'],ascending=False)
groupsB
```

| | المنطقة الإدارية | مجموع الطلبة | مجموع المعلمين | مجموع الإداريين |
|---|---|---|---|---|
| 4 | الشرقية | 1582.0 | 110.0 | 22.0 |
| 11 | مكة المكرمة | 1378.0 | 105.0 | 20.0 |
| 3 | الرياض | 1313.0 | 102.0 | 27.0 |
| 6 | المدينة المنورة | 1148.0 | 97.0 | 17.0 |
| 7 | تبوك | 1088.0 | 85.0 | 20.0 |
| 12 | نجران | 1054.0 | 73.0 | 16.0 |
| 8 | جازان | 956.0 | 71.0 | 17.0 |
| 1 | الجوف | 953.0 | 79.0 | 19.0 |
| 2 | الحدود الشمالية | 756.0 | 61.0 | 8.0 |
| 10 | عسير | 741.0 | 72.0 | 17.0 |
| 5 | القصيم | 708.0 | 72.0 | 15.0 |
| 9 | حائل | 648.0 | 69.0 | 15.0 |
| 0 | الباحة | 433.0 | 57.0 | 10.0 |

صنِّف البيانات بحسب الترتيب التنازلي.

الشكل 3.44: إنشاء مجموعة البيانات

لإنشاء مخطط شريطي أكثر وضوحًا، ستختار وتستخدم فقط الصفوف الخمسة الأولى من مجموعة البيانات الخاصة بك.

```
reg = groupsB.iloc[:5,0].tolist()
studentsH = groupsB.iloc[:5,1].tolist()
teacherH = groupsB.iloc[:5,2].tolist()
adminsH = groupsB.iloc[:5,3].tolist()
print(reg)
print (studentsH)
print(teacherH)
print(adminsH)
```

```
['تبوك' ,'المدينة المنورة' ,'الرياض' ,'مكة المكرمة' ,'الشرقية']
[1088.0 ,1148.0 ,1312.0 ,1378.0 ,1582.0]
[85.0 ,97.0 ,102.0 ,105.0 ,110.0]
[20.0 ,17.0 ,27.0 ,20.0 ,22.0]
```

الشكل 3.45: حدد الصفوف الخمسة الأولى من مجموعة البيانات

المقطع البرمجي لإنشاء المخطط الخاص بك.

```python
reg = groupsB.iloc[:5,0].tolist()
studentsH = groupsB.iloc[:5,1].tolist()
teacherH = groupsB.iloc[:5,2].tolist()
adminsH = groupsB.iloc[:5,3].tolist()
```

يمكنك استخدام هذا المقطع البرمجي، لعرض الأحرف العربية بشكل صحيح.

```python
regCor = [get_display(arabic_reshaper.reshape(element)) for element in reg]
x = np.arange(len(reg))  # the label locations
width = 0.5  # the width of the bars
```

إحداثيات الأعمدة.

```python
# This is a Matplotlib built-in style.
plt.style.use('fivethirtyeight')

# Determine the size of the figure
fig, ax = plt.subplots(figsize=(10,6))

rects1 = ax.bar(x, studentsH, width, label=get_display(arabic_reshaper.reshape('مجموع الطلبة')))

# Reshape Arabic letters to display them correctly
regionsLabel = get_display(arabic_reshaper.reshape('المناطق'))
meanLabel = get_display(arabic_reshaper.reshape('العدد المتوسط'))
title = get_display(arabic_reshaper.reshape('مجموع الطلبة والمعلمين والإداريين، أفضل 5 مناطق'))

# Set the labels to the axes and the title to the Bar chart
ax.set_xlabel(regionsLabel)
ax.set_ylabel(meanLabel)
ax.set_title(title)

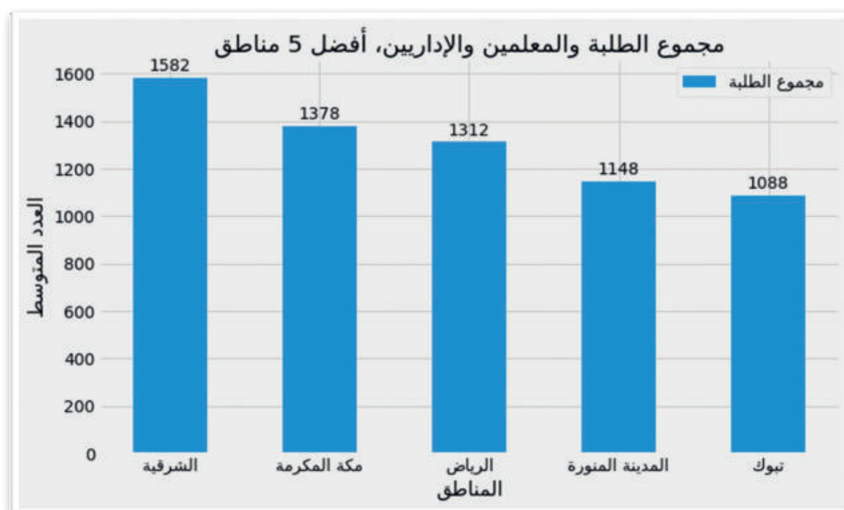
ax.set_xticks(x, regCor) # show the value under the bars
ax.legend() # show the legend

ax.bar_label(rects1, padding=3) # show each bar's value on the top of the bar
fig.tight_layout() # fit the chart in to the figure area
```

ضع التسمية على المخطط.

الشكل 3.46: إنشاء المخطط الشريطي

المخطط الشريطي الخاص بك جاهز.




الشكل 3.47: المخطط الشريطي



افترض الآن أنك تريد رسم عدد الطلبة والمعلمين والإداريين على نفس المخطط الشريطي. يسمى هذا المخطط بالمخطط الشريطي المجمّع وتحتاج إلى وضع الشرائط بشكل صحيح اعتمادًا على عرض الشريط.

```python
fig, ax = plt.subplots(figsize=(10,6))

studentsLabel = get_display(arabic_reshaper.reshape('مجموع الطلبة'))
teachersLabel = get_display(arabic_reshaper.reshape('مجموع المعلمين'))
adminsLabel = get_display(arabic_reshaper.reshape('مجموع الإداريين'))

rects1 = ax.bar(x - width/3, studentsH, width, label=studentsLabel)
rects2 = ax.bar(x, teacherH, width, label=teachersLabel)
rects3 = ax.bar(x + width/3, adminsH, width, label=adminsLabel)

# Add some text for labels, title and custom x axis tick labels, etc.

regionsLabel = get_display(arabic_reshaper.reshape('المناطق'))
meanLabel = get_display(arabic_reshaper.reshape('العدد المتوسط'))
title = get_display(arabic_reshaper.reshape('مجموع الطلبة والمعلمين والإداريين، أفضل 5 مناطق'))

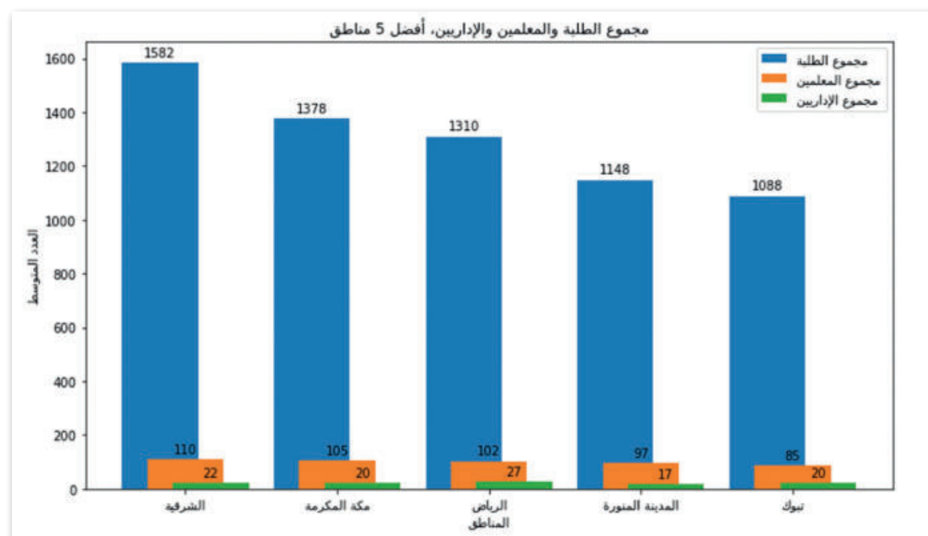
ax.set_xlabel(regionsLabel)
ax.set_ylabel(meanLabel)
ax.set_title(title)

ax.set_xticks(x, regCor)
ax.legend()

ax.bar_label(rects1, padding=3)
ax.bar_label(rects2, padding=3)
ax.bar_label(rects3, padding=3)


fig.tight_layout()
```

الشكل 3.48: إنشاء المخطط الشريطي المجمّع



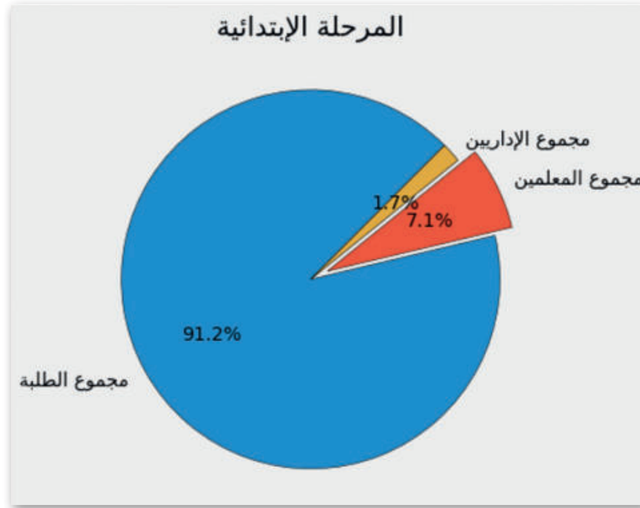

الشكل 3.49: المخطط الشريطي المجمّع



## المخطط الدائري Pie Chart

سترى خطوات إنشاء المخطط الدائري في مفكرة جوبيتر.

ستنشئ إطار بيانات جديد يسمى groupsP من مجموعة البيانات التي استخدمتها في الدرس السابق. قم بتجميع بياناتك حسب المرحلة واحصل على **المتوسط** (mean()) للطلبة والمعلمين والإداريين، ثم قم بفرز إطار البيانات هذا بمتوسط عدد المسؤولين.




الشكل 3.50:مخطط دائري

```
groupsP = data.groupby(['المرحلة'],as_index=False) [['مجموع الإداريين','مجموع المعلمين','مجموع الطلبة']].mean().round(0)
groupsP = groupsP.sort_values(by=['مجموع الإداريين'],ascending=False)
groupsP
```

| | المرحلة | مجموع الطلبة | مجموع المعلمين | مجموع الإداريين |
|---|---|---|---|---|
| 1 | المرحلة الإبتدائية | 1888.0 | 147.0 | 35.0 |
| 4 | رياض الأطفال | 575.0 | 52.0 | 25.0 |
| 3 | المرحلة المتوسطة | 870.0 | 72.0 | 15.0 |
| 2 | المرحلة الثانوية | 931.0 | 79.0 | 11.0 |
| 0 | التعليم المستمر | 119.0 | 0.0 | 1.0 |

الشكل 3.51: إنشاء إطار بيانات جديد

الآن ستنشئ مخطط دائري يوضح نسبة الطلبة والمعلمين والإداريين لمنطقة واحدة. أولاً، تحتاج إلى إنشاء قائمة تحتوي على شرائح المخطط الدائري. في مثالك، ستكون الشرائح عبارة عن قائمة تحتوي على أعداد الطلبة والمعلمين والإداريين للمنطقة في الصف الأول.

```
fig, ax = plt.subplots(figsize=(10,6), subplot_kw=dict(aspect="equal"))

#create the lists for the slices
slices = groupsP.iloc[0,1:].tolist()
labels = groupsP.columns[1:].tolist()
labelsCor = [get_display(arabic_reshaper.reshape(element)) for element in labels]

# array which specifies the fraction of the radius with which to offset each wedge
explode = [0,0.1,0]

# create the pie chart
ax.pie(slices, labels=labelsCor, explode=explode, shadow=False, startangle=45, autopct='%1.1f%%',
       wedgeprops={'edgecolor':'black'})

title = get_display(arabic_reshaper.reshape(groupsP.iloc[0,0]))
ax.set_title(title)
```

لإظهار النسبة المئوية لكل شريحة.

خصائص لمظهر المخطط الدائري.

الشكل 3.52: إنشاء مخطط دائري

الآن ستنشئ شكل يحتوي على أكثر من مخطط دائري.

```
fig, ([ax1,ax2],[ax3,ax4]) = plt.subplots(2,2,figsize=(16,10), subplot_kw=dict(aspect="equal"))

# First pie chart
slices = groupsP.iloc[0,1:].tolist()
labels = groupsP.columns[1:].tolist()
labelsCor = [get_display(arabic_reshaper.reshape(element)) for element in labels]

ax1.pie(slices, labels=labelsCor, shadow=False, startangle=45, autopct='%1.1f%%',
       wedgeprops={'edgecolor':'black'})

title1 = get_display(arabic_reshaper.reshape(groupsP.iloc[0,0]))
ax1.set_title(title1)

# Second pie chart
slices = groupsP.iloc[1,1:].tolist()
labels = groupsP.columns[1:].tolist()
labelsCor = [get_display(arabic_reshaper.reshape(element)) for element in labels]

ax2.pie(slices, labels=labelsCor, shadow=False, startangle=45, autopct='%1.1f%%',
       wedgeprops={'edgecolor':'black'})

title2 = get_display(arabic_reshaper.reshape(groupsP.iloc[1,0]))
ax2.set_title(title2)

# Third pie chart
slices = groupsP.iloc[2,1:].tolist()
labels = groupsP.columns[1:].tolist()
labelsCor = [get_display(arabic_reshaper.reshape(element)) for element in labels]

ax3.pie(slices, labels=labelsCor, shadow=False, startangle=45, autopct='%1.1f%%',
       wedgeprops={'edgecolor':'black'})

title3 = get_display(arabic_reshaper.reshape(groupsP.iloc[2,0]))
ax3.set_title(title3)
# Fourth pie chart
slices = groupsP.iloc[3,1:].tolist()
labels = groupsP.columns[1:].tolist()
labelsCor = [get_display(arabic_reshaper.reshape(element)) for element in labels]

ax4.pie(slices, labels=labelsCor, shadow=False, startangle=45, autopct='%1.1f%%',
       wedgeprops={'edgecolor':'black'})

title4 = get_display(arabic_reshaper.reshape(groupsP.iloc[3,0]))
ax4.set_title(title4)
```

الشكل 3.53: إنشاء أربعة مخططات دائرية

**تحتوي مكتبة مات بلوت ليب على مجموعة كبيرة من الإعدادات لتصميم مخطط أو رسم بياني. يمكنك زيارة موقع المكتبة الإلكتروني http:\\matplotlib.org والاطلاع على دليل المستخدم للحصول على تفاصيل كافة إعداداتها.**



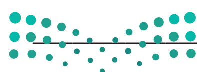



الشكل 3.54: المخطط الدائري

# تمرينات

**1** ناقش أهمية تصوير البيانات كمرحلة في التحليل الاستكشافي للبيانات، ثم وضّح أهميته بمثالين.

**2** قارن بين الخصائص الرئيسة للمخططات البيانية الخطّية والشريطية، ثم اذكر مثالين لمجموعات البيانات واختر المخطط المناسب لكل مثال.



**3** حدد الفرق الرئيسي بين المخطط النقطي والمخططات الأخرى، ثم اذكر مثالًا على استخدام المخطط النقطي.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**4** ادرج مكتبات بايثون التي تحتاجها لتطبيق تقنيات تصوير البيانات، وما هي الخطوات لبدء استخدامها في مفكرة جوبيتر؟

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**5** تريد معرفة عدد السياح الذين يزورون المملكة العربية السعودية شهريًا لمنطقة واحدة في مجموعة البيانات.

- ما نوع المخطط الأنسب للاستخدام؟ (ادعم إجابتك بأسباب منطقية).
- اختر أي منطقة من مجموعة البيانات واستخدم مكتبة مات بلوت ليب، وأنشئ نوع المخطط الذي تعتقد أنه الأنسب.
- بناءً على المخطط الذي أنشأته، اكتشف الشهر الذي استقبلت فيه المنطقة التي اخترتها أكبر عدد من الزوار.



**6** تريد مقارنة عدد السياح الذين يزورون المملكة العربية السعودية من 3 أجزاء من العالم، أوروبا وآسيا والشرق الأوسط من شهر أكتوبر إلى يناير.

- ما نوع المخطط الأنسب للاستخدام؟ (ادعم إجابتك بأسباب منطقية).
- أنشئ نوع المخطط الذي تعتقد أنه الأنسب.
- بناءً على المخطط الذي أنشأته، اكتشف أي جزء من العالم جاء منه معظم السياح في كل شهر من أكتوبر إلى يناير.

**7** تريد معرفة الشهر الذي استقطب معظم الزوار ثم معرفة النسبة المئوية لزوار كل منطقة لهذا الشهر.

- أنشئ إطار بيانات لتحديد الشهر الذي يضم أكبر عدد من الزوار.
- أي نوع من المخططات سيكون الأنسب في هذه الحالة؟ (ادعم إجابتك بأسباب منطقية).
- أنشئ نوع المخطط الذي تعتقد أنه الأنسب.
- بناءً على المخطط الذي أنشأته، ما هي المنطقة التي لديها أعلى نسبة من الزوار وأي منها لديها أقل نسبة من الزوار؟



# المشروع

تريد معرفة الطريقة الأكثر تفضيلاً للسياح لزيارة المملكة العربية السعودية. لديك ملف إكسل بالاسم "tourist-indicators.xlsx" والذي يحتوي على معلومات حول عدد السياح الذين يصلون إلى المملكة العربية السعودية عن طريق الجو والبر والبحر شهريًا.

1. افتح الملف "tourist-indicators.xlsx".

2. قم بتحميل ورقة العمل "I7" في إطار بيانات جديد باستخدام مكتبة بانداس.

3. احصر عدد السياح الذين يصلون عن طريق الجو والبر والبحر لكل شهر.

4. قارن متوسط عدد السياح الذين يصلون إلى المملكة العربية السعودية عن طريق الجو والبر لأشهر يناير وفبراير ومارس باستخدام تقنية التصوير المناسبة.

5. ما هي النسب المئوية لكل طريقة وصول للأشهر الثلاثة مع أقل عدد إجمالي من الزوار؟ للإجابة على السؤال، تحتاج إلى إنشاء عمود جديد في إطار البيانات الخاص بك مع إجمالي عدد الزوار شهريًا.

6. ما نوع المخطط الذي سيكون أكثر فائدة للإجابة على هذا السؤال؟ ادعم إجابتك بأسباب منطقية.



# ماذا تعلمت

- خطوات عملية تحليل البيانات.
- استخدام مفكرة جوبيتر كأداة لتحليل البيانات.
- استخدام مكتبة بانداس لإنشاء الإحصاءات.
- أهمية تصوير البيانات.
- استخدام مكتبة مات بلوت ليب لتمثيل البيانات بشكل رسومي.
- إنشاء مخطط بياني شريطي ومخطط دائري في مفكرة جوبيتر.

## المصطلحات الرئيسة

| | | | |
|---|---|---|---|
| Attribute | سمة | Grouping | تجميع |
| Data Cleaning | تنظيف البيانات | Indexing | فهرسة |
| Data Frame | إطار البيانات | Method | وظيفة |
| Data Visualization | تصوير البيانات | Multivariate | متعدد المتغيرات |
| Descriptive Analysis | التحليل الوصفي | Non-Graphical Analysis | تحليل غير رسومي |
| Diagnostic Analysis | التحليل التشخيصي | Predictive Analysis | التحليل التنبؤي |
| Exploratory Data Analysis | تحليل البيانات الاستكشافي | Prescriptive Analysis | التحليل التوجيهي |
| Filtering | تصفية | Programming Library | مكتبة البرمجة |
| Function | دالة | Series Object | كائن المتسلسلة |
| Graphical Analysis | تحليل رسومي | Univariate | أحادي المتغير |



# 4. نمذجة البيانات التنبؤية والتوقُّع

سيتعرّف الطالب في هذه الوحدة على مفهوم نمذجة البيانات التنبؤية والتوقُّع، ويشمل ذلك مفهوم النمذجة التنبؤية، وأنماط النماذج التنبؤية وتطبيقاتها. وأيضًا، سيتعلم الطالب مفهوم التوقُّع، وشرح وتوضيح خطوات الوصول إلى نتائج التوقُّع، وسيتم التركيز على مفهوم تحسين الحلول، وذلك من خلال صياغة المشكلة وإيجاد الحل الأمثل لها من بين الحلول المُمْكنة باستخدام أداة إكسل سولفر (Excel Solver).

وفي الختام، سيتعلّم الطالب طريقة تقييم النتائج التي يحصل عليها وذلك بهدف الوصول إلى أفضل النتائج والتوصيات المتعلقة بالإجراءات أو العمليات التي قد يتم تنفيذها في المستقبل.

**بنهاية هذه الوحدة سيكون الطالب قادرًا على أن:**

- يُعرِّف مصطلح النمذجة التنبؤية.
- يُقدِّم وصفًا واضحًا لفئات النمذجة التنبؤية.
- يُحدِّد خطوات عملية النمذجة التنبؤية.
- يُعدِّد ميزات وعيوب النمذجة التنبؤية.
- يُعرِّف مصطلح التوقُّع.
- يُحدد الخطوات المتبعة في عمليات التنبؤ.
- يُنفِّذ عملية التوقُّع في مايكروسوفت إكسل (Microsoft Excel).
- يُعَرِّف مفهوم نطاق الثقة.
- يُصنِّف المخططات أو الرسومات البيانية المتنوعة للتوقُّع.
- يُعرِّف مفهوم نموذج التحسين.
- يُطبِّق عملية التَّحسين باستخدام أداة إكسل سولفر.
- يُقيِّم نتائج عملية التحسين ويُحدِّد التوصيات المستقبلية.





رابط الدرس الرقمي

www.ien.edu.sa

الدرس الأول

# نمذجة البيانات التنبؤية

تلجأ المؤسسات والشركات لاستخدام النمذجة التنبؤية لتحليل الأحداث المستقبلية المتعلقة بنشاطها التجاري، وذلك بهدَف اتخاذ أفضل القرارات. ويُمكِن استخدام نماذج التّنبؤ لفهم ومعرفة شرائح وفئات المستهلكين، ولتقدير المبيعات المُحتملة، أو لفهم ومعرفة القضايا الأمنية للحسابات.

## ما هي النمذجة التنبؤية؟ What is Predictive Modeling?

تُعتبر التحليلات التنبؤية فرعًا من فروع علم تحليل البيانات المتقدم، وتَستعين هذه التحليلات بالبيانات السابقة، إلى جانب طُرق أخرى كالنمذجة الإحصائية، وتنقيب البيانات، وتعلُّم الآلة، وذلك لتقديم التنبؤات حول النتائج المُستقبلية لقرارات أو لعمليات مُعينة تقوم بها الشركات أو المؤسسات. وتستخدم الشركات والمؤسسات التحليلات التنبؤية للتعرُّف على أنماط مُعينة في هذه البيانات يُمكن من خلالها تحديد الفُرص والمخاطر. فعلى سبيل المثال، تَجمعُ خدمة الأرصاد الجوية البيانات بشكل يومي عن المتغيرات المختلفة المتعلقة بحالة الطقس مثل درجات الحرارة والرطوبة وغيرها، مما يُمكِّنها من التنبُّؤ بحالة الطقس في الأيام القادمة.

تُستَخدم التحليلات التنبؤية على نطاقٍ واسع في مجال الرعاية الصحية وذلك بِهدف تحسين طُرق تشخيص وعلاج المرضى المصابين بالأمراض المُزمنة، وتستخدم إدارات الموارد البشرية والشركات نماذج التنبؤ في تحسين عمليات اختيار وتعيين الموظفين، وأما البنوك فتستخدمها بشكلٍ واسع للكشف عن عمليات الاحتيال.

**النمذجة التنبؤية (Predictive Modeling):**

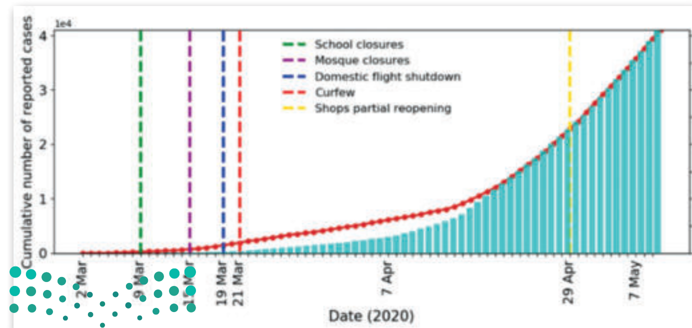
هي أسلوبٌ إحصائي تُستَخدمُ فيه النتائج والبيانات السابقة للتنبؤ بالأحداث أو النتائج المستقبلية.

### مثال

عندما أصبح **فيروس كورونا** (COVID-19) وباءً وأصاب جميع الدول في أنحاء العالم، اعتمد خبراء الصحة في كثير من الدول على علم البيانات لنمذجة السلوك الوبائي للمرض وللتنبؤ بمعدلات العدوى والوفيات. ولقد أسهمت النماذج التي تم تطويرها في تمكين الجهات الصحية والباحثين والعاملين في المجال الطبي من تطوير طرقٍ لكبح جماح انتشار الجائحة، والحدِّ من آثارها المحتملة.

أجرى باحثون في جامعة الملك سعود في المملكة العربية السعودية –وذلك بالتعاون مع جامعات أخرى– دراسةً تنبؤية حول انتشار **فيروس كورونا** (COVID-19) في المملكة، وهَدَفت تلك الدِّراسة إلى التوصل إلى فهم عميق للسلوك المتغير للعدوى باستخدام النماذج التنبؤية والمحاكاة، واستعان الباحثون ببيانات واحصائيات دقيقة صادرة من وزارة الصحة السعودية لدعم نماذجهم بمعلومات عن انتشار الوباء، ولتقديم التوقُّعات المحتملة حول أعداد الإصابات. لقد ساعد هذا التقدير في اتخاذ القرارات المناسبة من قِبَل الحكومة والجهات المسؤولة في المملكة، وذلك من خلال اتخاذ التدابير الفعالة للمراقبة والوقاية، وتضمَّنت هذه التدابير فرض القيود على السفر والتنقل وإغلاق المدارس والمساجد، وكان لها تأثيرٌ عظيمٌ في تأخير الوصول لذروة تفشي الوباء والحدِّ من مُعدَّلات الإصابة وانتشار الوباء في المملكة.

لقد أصبح القيام بعمليات النمذجة أمرًا مهمًا خلال تلك الفترة، ويرجِع ذلك بشكلٍ أساس إلى توفُّر البيانات الحقيقية، التي ساهمت في تقديم نماذج التنبؤ لانتشار المرض من خلال مقارنة الأعداد المتوقّعة للإصابات بالعدد الفعلي لها. أصبح عدد الحالات المكتشفة حديثًا يتناقص مع دخول الإجراءات المختلفة مثل الإغلاق وقيود السفر حيِّز التنفيذ، وكان في ذلك دليلٌ واضحٌ على أن تنبؤات الباحثين كانت قريبةً جدًا مما حدث فعلًا، كما يظهر في الشكل 4.1، حيث تُظهر الاعمدة في الشكل المعلومات التراكمية حول الأعداد الحقيقية للإصابة، بينما يُظهر الخط الأحمر توقُّعات عدد الإصابات، ويَعرض المخطط البياني أيضًا التواريخ التي تم فيها فرض القيود المختلفة.




شكل 4.1: تقييم النموذج التنبؤي مع العدد الفعلي والعدد المُتوقَّع لحالات الإصابة المسجلة يوميًا

## فئات النمذجة التنبؤية Predictive Modeling Categories

تتمثل مُهمة المُتعلِّم في النمذجة التنبؤية بالوصول إلى الدالة أو العلاقة الوظيفية التي تربط متغيرات الإدخال بالمخرجات (التنبؤات) في **بيانات التدريب** (Training Data)، وذلك بصرف النظر عن طبيعة ومُعاملات تلك الدالة.

بمجرد الوصول إلى هذه العلاقة الوظيفية، يُمكن استخدامها للتنبؤ بقيَم المُخرجات بناءً على متغيرات الإدخال المختلفة. وتصنف النماذج التنبؤية إلى فئتين: فئة تحتوي على عدد محدد من المُعاملات وتسمى **بالنموذج المُعاملي** (Parametric Model)، وفئة لا تحتوي على عدد محدد من المُعاملات، ويطلق عليها تسمية **النموذج غير المُعاملي** (Non-Parametric Model).

### 1. النماذج المُعاملية Parametric Models

تعتبر الافتراضات جزءًا أساسيًا من أي نموذج من نماذج البيانات، فهي تُحسّن التنبؤات وتجعل النموذج أسهل للفهم. يَضعُ النموذج المُعاملي افتراضات محددة حول شكل الدالة التي سيتم تعيينها، ويفترض مجموعة محددة مُسبقًا من المُعاملات، وذلك بشكل مستقل عن تلك الموجودة في أمثلة التدريب، وهكذا فإن النموذج المُعاملي يقوم بتلخيص بيانات التدريب من خلال هذه المجموعة من المُعاملات.

**المُعامل (Parameter):**
يمكن وصف المُعامل بأنه متغير جوهري وأساسي في تكوين النموذج.

### 2. النماذج غير المُعاملية Non-Parametric Models

إن نماذج تَعلُّم الآلة غير المُعاملية ليست مَعنيَّة بتكوين الافتراضات حول **دالة التعيين** (Mapping Function)، فيمكن لمثل هذه النماذج مثلًا تقدير طبيعة العلاقة الوظيفية من خلال بيانات التدريب. وتُعدُّ هذه النماذج خيارًا ممتازًا لتحليل الكميات الكبيرة من البيانات بدون أي معرفة سابقة عنها.

**يعتمد المتخصصون في عمل تحليلات النماذج التنبؤية على البيانات من المصادر التالية:**

| |
|---|
| بيانات عملية (Transactional Data). |
| بيانات العملاء (Customer Data). |
| البيانات الطبية (Medical Data). |
| البيانات المالية (Financial Data). |
| المعلومات الديموغرافية (Demographic Data). |
| البيانات الجغرافية (Geographic Data). |
| بيانات التسويق الرقمي (Digital Marketing Data ). |
| إحصائيات الويب (Web Traffic Statistics). |

**الجدول 4.1: مقارنة بين النماذج المُعاملية وغير المُعاملية**

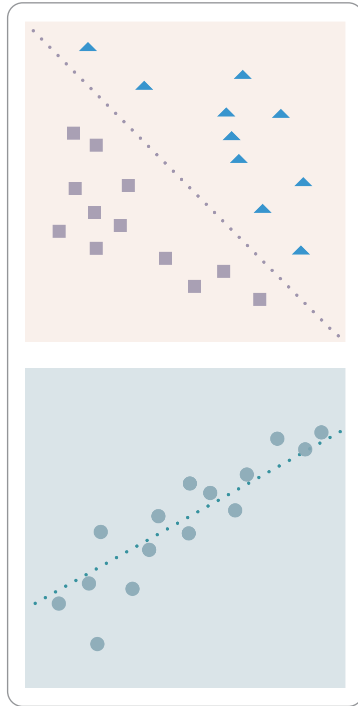
| المعيار | النماذج المُعاملية | النماذج غير المُعاملية |
|---|---|---|
| بيانات التدريب | تتطلب بيانات تدريب أقل من النماذج غير المُعاملية. | تتطلب بيانات أكثر بكثير من النماذج المُعاملية لتقدير العلاقة أو دالة التعيين. |
| سرعة التدريب | أسرع إنجازًا من الناحية الحسابية، ويُمكن تدريبها بشكل أسرع لوجود مُعاملات محدودة للتدريب. | تستغرق وقتًا أطول للتدريب، حيث تتضمن تحليل علاقات أكثر تعقيدًا يتم تقديرها أثناء عملية التدريب. |
| الملاءمة | قد لا تُقدِّم هذه النماذج أفضل ملاءمة للبيانات، ومن المستبعد أن تتطابق تمامًا مع دالة التعيين. | تُوفِّر هذه النماذج تنبؤات أكثر دقة من النماذج المُعاملية من حيث ملاءمة البيانات، ولكن الخوارزميات في هذه النماذج تكون أكثر عرضة لمشكلة فرط التخصيص (Overfitting). |
| التعقيد | تتميز إجراءاتها بسهولة فهمها وتفسيرها. | إجراءاتها أكثر تعقيدًا وصعوبة سواء من ناحية إمكانية التفسير أو الفهم. |

## مهام النمذجة التنبؤية Predictive Modeling Tasks

تُعدُّ نماذج **التصنيف** (Classification) و**الانحدار** (Regression) من أهم وأكثر النماذج استخدامًا في مهام النمذجة التنبؤية.

### 1. التصنيف Classification

يعتمد نموذج التصنيف على عملية تقييم المتغيرات المُدخلة ثم تصنيفها ضمن مجموعات لتكوين بيانات المُخرجات، وبذلك فإن المتغير الذي سيتم توقُّعه ستكون له قيمًا **متقطّعة** (Discrete)، وقد تكون هذه القيم ببساطة مجرد إجابة لسؤال معين بـ "نعم" أو "لا". ويُستخدم نموذج التصنيف في تقييم عمليات التمويل والبيع بالتجزئة، حيث بمقدوره جمع المعلومات بسرعة وتصنيفها في مجموعات لتقديم الإجابات عن الأسئلة المتعلقة بتلك العمليات.



شكل 4.2: يوضح الفرق بين التصنيف (الشكل العلوي) والانحدار (الشكل السفلي)، حيث يمثّل التصنيف الخط المُنقط وهو الحد الخطي الفاصل بين فئتين مختلفتين، بينما يُمثل الخط المُنقط في الانحدار العلاقة الخطية بين متغيرين.

### 2. الانحدار Regression

يعتمد نموذج الانحدار على مبدأ إيجاد علاقات رياضية تربط بين متغيرين، بحيث يُمكن تنبؤ أحدهما من خلال معرفة المتغير الآخر، ويُطلق على المتغير المُدخل اسم **المتغير المستقل** (Independent Variable)، بينما يُطلق على المتغير المُخرج اسم **المتغير التابع** (Dependent Variable)، ويتنبأ هذا النموذج بالقيم المحتملة للمتغيرات التابعة من خلال معالجة قيم المتغيرات المستقلة. يتم تمثيل هذا النموذج بيانيًا في صورة خط مستقيم (انحدار خطي) يتقارب مع جميع نقاط البيانات المستقلة. ويمكن لنموذج الانحدار على سبيل المثال التنبؤ بمدة بقاء شخص إبان دخول المستشفى، ويمثل عدد الأيام في المستشفى المتغير التابع، أما معدل النبض لذلك الشخص مثلاً فيمثل المتغير المستقل.

**الجدول 4.2: مقارنة بين التصنيف والانحدار**

| التصنيف | الانحدار |
|---|---|
| التصنيف هو التنبؤ بالمُخرجات لفئة متقطعة بمعنى أن المتغير المُخرج يجب أن يكون عددًا صحيحًا. | الانحدار هو التنبؤ بناتج كمي مستمر بمعنى أن المتغير المُخرج يجب أن يكون قيمة مستمرة أو عددًا حقيقيًا. |
| تُستخدم خوارزمية التصنيف لتعيين قيمة المُدخل (x) مع المتغير المُخرج ذو القيم المتقطّعة (y). | تُستخدم خوارزمية الانحدار لتعيين قيمة المُدخل (x) مع المتغير المُخرج ذو القيم المستمرة (y). |



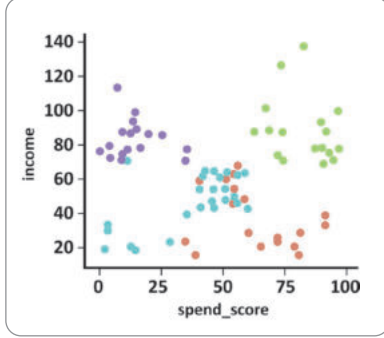
من المهام الشائعة الأخرى للنمذجة التنبؤية:

### 3. التوقُّع Forecasting

وهو إجراء وتقديم تقديرات رقمية معينة بناءً على تحليل البيانات السابقة والتي يطلق عليها البيانات التاريخية. وتَستخدم شركات الاستثمار التوقُّعات للتنبؤ بأسعار الأسهم في التداولات اليومية أو طويلة الأجل، ويعتبر نموذج التوقُّع من أكثر نماذج التنبؤ شيوعًا حيث يتميز بامكانيات استخدام كثيرة في العديد من المجالات.

### 4. التجميع Clustering

يُصنِّف نموذج التجميع البيانات إلى مجموعات بناءً على الخصائص المتشابهة بينها، ثم يستخدم بيانات كل **مجموعة** (Cluster) لتحديد النتائج على نطاق واسع لكل مجموعة. وهناك نوعان من طرق التجميع يتم استخدامهما في هذا النموذج: **التجميع الصلب** (Hard Clustering) يعتمد على تصنيف البيانات إلى مجموعات متميزة، حيث يمكن أن تنتمي كل نقطة بيانات إلى مجموعة واحدة فقط، و**التجميع الناعم** (Soft Clustering) يعتمد على تعيين احتمالات لكل نقطة بيانات، حيث يمكن أن تنتمي نقاط البيانات إلى أكثر من مجموعة واحدة. ويُمكن للشركات استخدام نموذج التجميع لتحديد استراتيجيات التسويق لفئات معينة من المستهلكين.



شكل 4.3: مثال على التجميع لأربع مجموعات بناء على قيمتي الدخل ومعدل الإنفاق

### 5. نموذج اكتشاف القيم الشاذة Outlier Detection

يطلق مصطلح القيم الشاذة على قيم البيانات غير المتجانسة أو تلك المنفصلة عن بقية البيانات في مجموعة البيانات، ويمكن لنماذج اكتشاف القيم الشاذة فحص وتحديد القيم الغريبة وغير العادية في البيانات، وتقييم مدى ارتباطها بفئات أو أرقام أخرى.

### 6. السلاسل الزمنية Time Series

تَستخدِم نماذج السلاسل الزمنية قيمَ البيانات المتوفرة سابقًا ضمن تسلسل زمني مُحدد كعوامل الإدخال في مجموعة البيانات؛ وذلك من أجل التنبؤ بقيم جديدة أو أحداث مستقبلية، ويمكن لهذه النماذج تقديم التوقُّعات المستقبلية لاتجاهات أو أحداث فريدة أو متعددة. يمكن لنماذج السّلاسل الزمنية أيضًا تحليل تأثير العوامل الخارجية كتلك الموسمية والعارضة (غير المتوقّعة) التي قد تحدث على القيم والاتجاهات المستقبلية، على سبيل المثال يمكن لشركة صناعات إلكترونية استخدام نموذج السلاسل الزمنية لتحليل الوقت المطلوب لمعالجة الطلبيات على مدار العام الماضي، وبالتالي يمكن للنموذج التنبؤ بمتوسط وقت المعالجة الشهري.

تُستخدم طرق أخرى للنمذجة التنبؤية في المسائل الأكثر تعقيدًا.

| من طُرق النمذجة التنبؤية: |
| --- |
| أشجار القرار (Decision Trees). |
| التعزيز الاشتقاقي (Gradient Boosting). |
| النماذج الخطية العامة (General linear Models). |
| الشبكات العصبية (Neural Networks). |
| نماذج بروفيت (Prophet Models). |



## عملية النمذجة التنبؤية The Predictive Modeling Process

يمكن تعريف النمذجة التنبؤية ببساطة على أنها عملية تنفيذ خوارزميات على مجموعات من البيانات لإنشاء التنبؤات، ويتم في هذه العملية إنشاء نموذج وتدريبه، ثم التحقق من صِحته وإدخال التحسينات عليه عند الحاجة، للحصول على المعلومات المناسبة التي تُلبي احتياجات المؤسسة. وتتكون الخطوات الأساسية لإجراء النمذجة التنبؤية بشكل نموذجي من:

### 1. جمع البيانات وتنظيفها Data collection and cleaning

إن من المهم القيام بجمع البيانات من جميع المصادر المتوفرة بهدف استخراج المعلومات اللازمة لعملية النمذجة، وبعد ذلك تتم عملية تنظيفها من الشوائب والقيم الشاذة للحصول على تقديرات دقيقة. وتُطبَّق هذه الخطوة على: البيانات المختلفة مثل عمليات البيع والشراء والاستبانات الخاصة بالعملاء، والبيانات الإحصائية الخاصة بالاقتصاد والمسح السكاني، والبيانات التي يتم جمعها بشكلٍ آلي عبر الويب ومن خلال الأجهزة المختلفة وغير ذلك.

### 2. تحويل البيانات Data transformation

تتم عملية تحويل البيانات بتوحيد بُنية وصياغة البيانات باستخدام عمليات معالجة دقيقة للحصول على البيانات في صورتها النهائية، وتشمل هذه العملية تحديد نطاقات معينة لقيم البيانات وإزالة القيم الغريبة والبيانات الشاذة من خلال **تحليل الارتباط** (Correlation Analysis).

### 3. صياغة النموذج التنبؤي Formulation of the Predictive Model

تتضمن عملية صياغة النموذج التنبؤي القيام بتحديد طرق التنبؤ المناسبة حسب الحاجة، فيمكن مثلًا استخدام شجرة القرار في عملية التصنيف، بينما يجب استخدام نموذج التعزيز الاشتقاقي حين تكون المهمة تتعلق بالانحدار. ويتم أثناء هذه العملية تحديد بيانات التدريب والاختبار في النموذج، حيث يتم تدريب خوارزمية الإجراء المحدد باستخدام بيانات التدريب المتاحة، ثم يتم تطبيق النموذج الناتج على البيانات لاختبارها وتحديد أداء النموذج.

### 4. الاستنتاجات أو الاستدلالات Inferences or conclusions

في النهاية يتم استخراج الاستدلالات واستخلاص الاستنتاجات من النموذج، والتي تُساعد في الإجابة على أسئلة الأعمال.




شكل 4.4: مخطط عملية النمذجة التنبؤية

## مثال عملي على التصنيف Practical Classification Example

يعرض هذا المثال كيفية إنشاء نموذج تنبؤي في إطار علم البيانات. افترض أنك تعمل ضمن فريق مشروع مهمته تفقُّد هياكل المباني الخرسانية وفحصها بحثًا عن الشقوق الموجودة فيها، ولأن هذه العملية تتميز بالصعوبة والخطورة المصاحبة لها، وبشكلٍ خاص في المباني المرتفعة، فيمكنك إنشاء **نموذج تَعلُّم الآلة** (Machine Learning Model) الذي بمقدوره فحص صور الخرسانة في هياكل المباني، ثم تصنيفها إلى فئةٍ إيجابية في حال وجود الشقوق بها، وأخرى سلبية إذا خلَتْ من الشقوق. يمكن بعد ذلك دمج الصور التي يمكن التقاطها بواسطة طائرة مسيَّرة بدون طيار مع النموذج مما يتيحُ إجراء فحص المباني بطريقة أكثر أمانًا وفاعلية.

تتطلب عملية تدريب النموذج وجود البيانات، والتي ستنقسم في هذه الحالة إلى فئتين: تتَضمن الفئة الأولى صورًا للخرسانة التي تحتوي على الشقوق، بينما تتضمن الأخرى صورًا لخرسانة سليمةٍ خالية من الشقوق.

يجب أيضًا تقسيم مجموعة بيانات الصور إلى مجموعتي بيانات منفصلتين:

- **مجموعة التدريب** (Training Dataset) وتتضمن الصور التي ستستخدمها لتدريب نموذج تعلُّم الآلة.
- **مجموعة الاختبار** (Test Dataset) وتتضمن صورًا جديدة لم تكن ضمن مجموعة بيانات تدريب النموذج ويهدف استخدام هذه الصور إلى اختبار أداء النموذج وتقييمه.

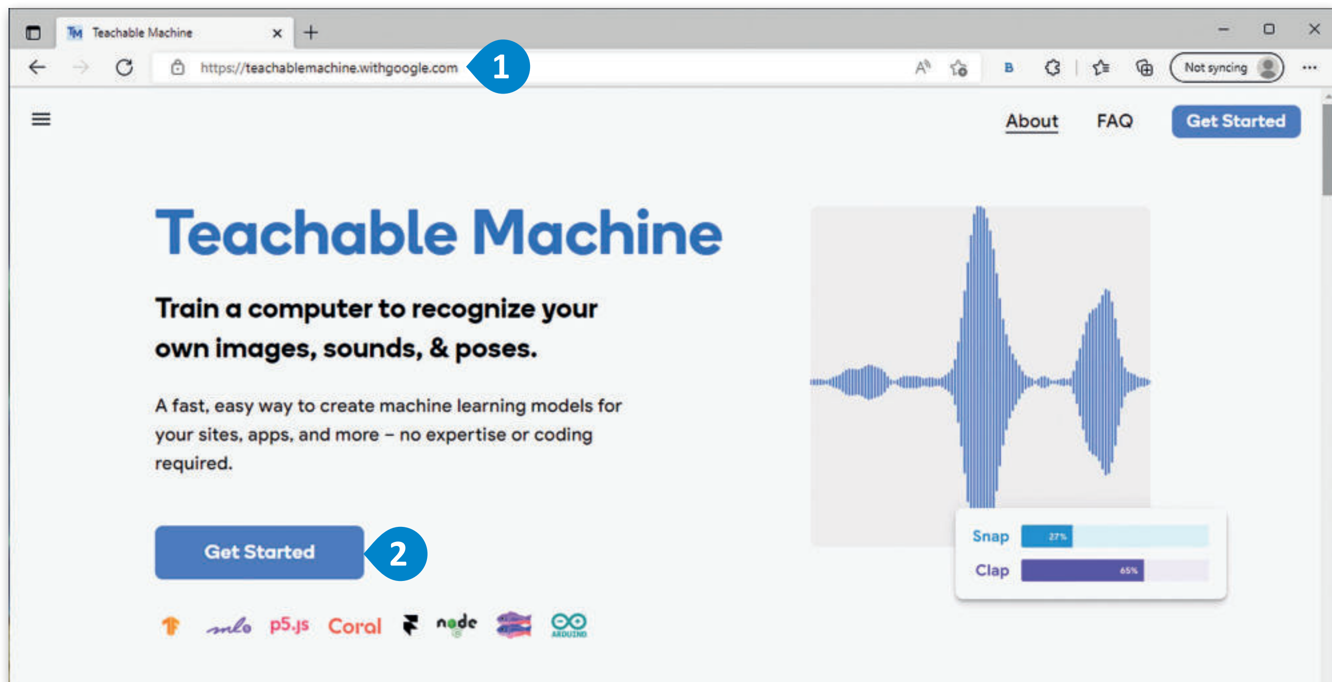
يجب أن تحتوي مجموعة التدريب ومجموعة الاختبار على صورٍ لكلتا الفئتين من الهياكل الخرسانية: الفئة الإيجابية (التي تحتوي على الشقوق) أو من الفئة السلبية (التي تخلو من الشقوق).

لتدريب نموذج على تصنيف صور الخرسانة، يمكنك استخدام برنامج آلة قابلة للتعليم (Teachable Machine) عبر الإنترنت، وهي مُتاحة على الموقع الإلكتروني: https://teachablemachine.withgoogle.com، وستقوم بتحميل الصور من مجلد الصور في حاسوبك لتصنيفها.

### لإنشاء نموذج وتدريبه:

- افتح المُتصفح وانتقل إلى الموقع الإلكتروني: https://teachablemachine.withgoogle.com. 1
- اضغط على Get Started (البدء). 2
- اضغط على Image Project (مشروع الصورة). 3
- اضغط على Standard Image Model (نَموذج الصورة القياسي). 4
- أعد تسمية Class 1 (الفئة1) إلى Positive (إيجابي) و Class 2 (الفئة 2) إلى Negative (سلبي). 5
- اضغط على Upload (تَحميل) في الفئة الإيجابية. 6
- اضغط على Choose images from your files, or drag & drop here (اختيار الصور من ملفاتك أو اسحبها وأفلتها هنا)، 7 لتحديد وتحميل مجموعة الصور التدريبية التي بها شقوق في الخرسانة من المجلد الفرعي **إيجابي** (Positive) في مجلد **صور التصنيف** (Images for classification) الموجود داخل **المستندات** (Documents).
- كَرِّر العملية لتحديد وتحميل مجموعة الصور التدريبية التي لا تحتوي على شقوق في الخرسانة من المجلد الفرعي **سلبي** (Negative) في مجلد **صور التصنيف** (Images for classification) داخل **المستندات** (Documents). 8
- اضغط على Train Model (تدريب النموذج). 9






Teachable Machine

https://teachablemachine.withgoogle.com/train

Not syncing

Teachable Machine

# New Project

Open an existing project from Drive.

Open an existing project from a file.

## Image Project

Teach based on images, from files or your webcam.

## Audio Project

Teach based on one-second-long sounds, from files or your microphone.

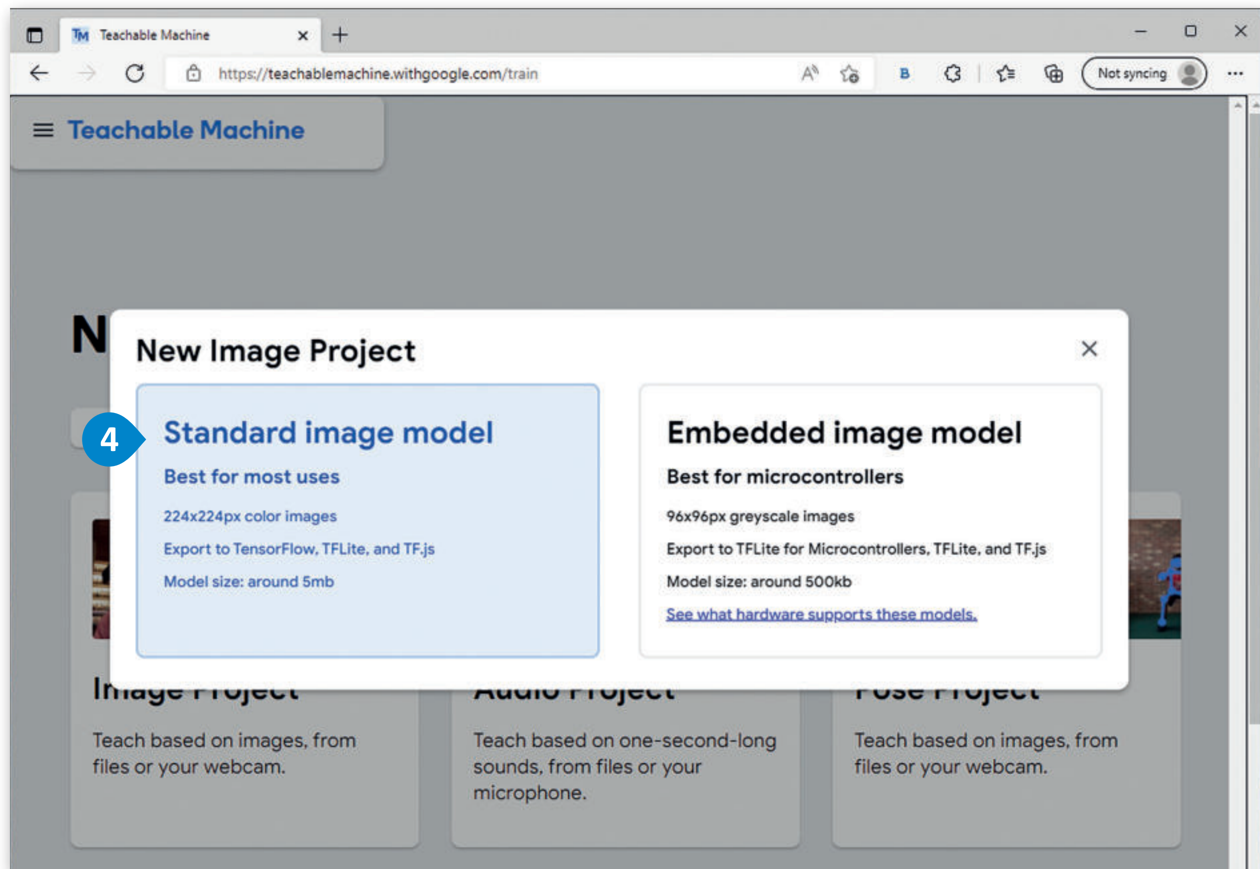
## Pose Project

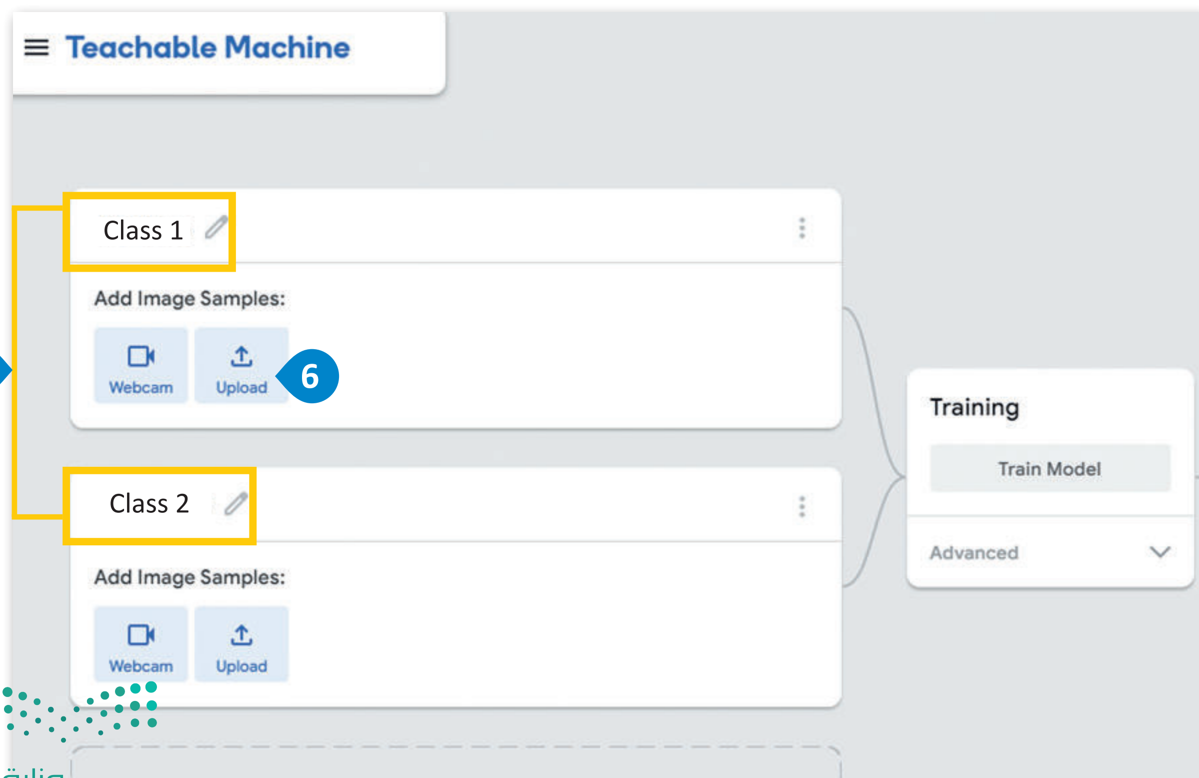
Teach based on images, from files or your webcam.

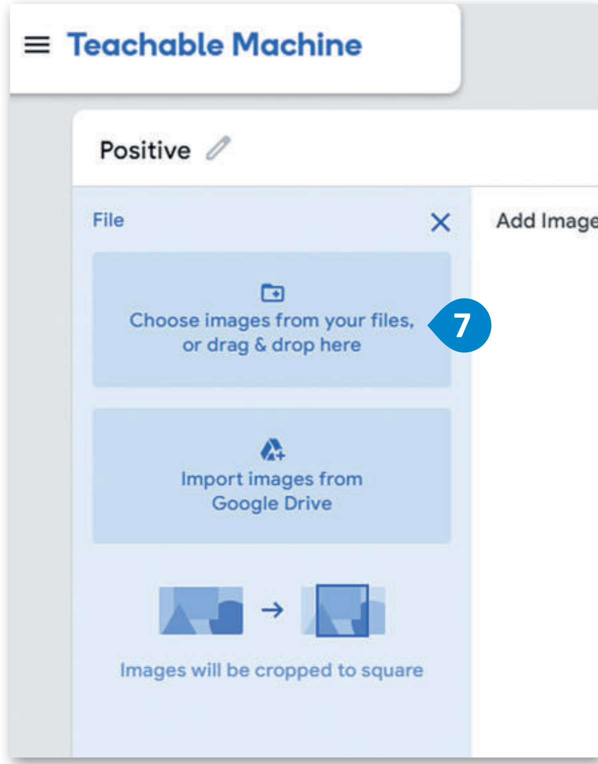
3

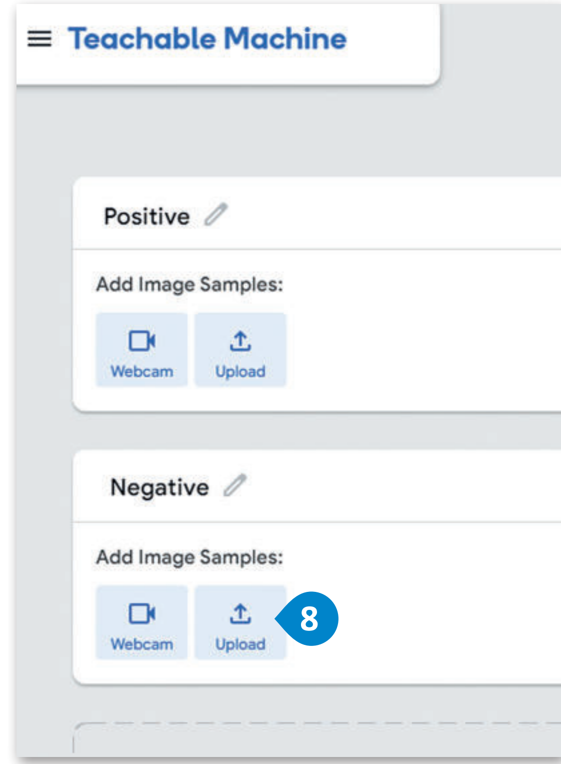
English

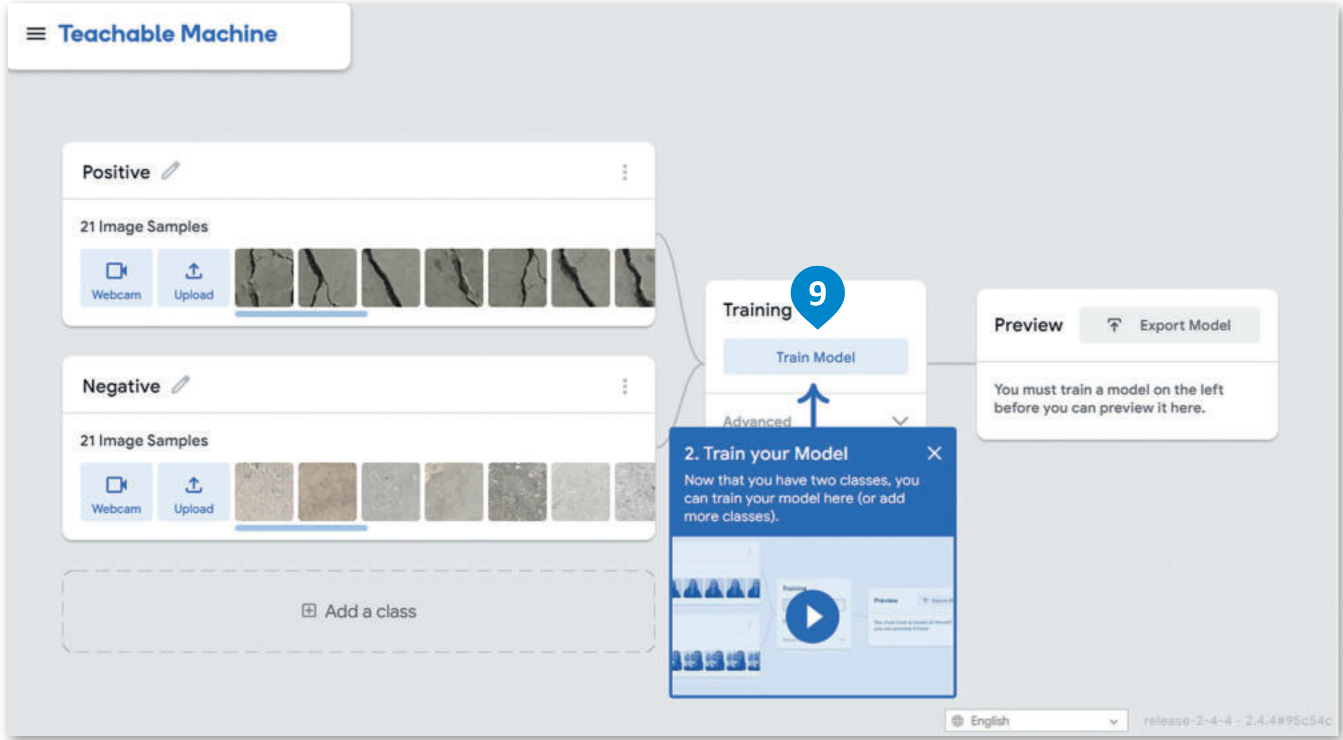
release-2-4-4 - 2.4.4#95c54c


Teachable Machine
https://teachablemachine.withgoogle.com/train
Not syncing
Teachable Machine
New Image Project
4
Standard image model
Best for most uses
224x224px color images
Export to TensorFlow, TFLite, and TF.js
Model size: around 5mb
Embedded image model
Best for microcontrollers
96x96px greyscale images
Export to TFLite for Microcontrollers, TFLite, and TF.js
Model size: around 500kb
See what hardware supports these models.
Teach based on images, from files or your webcam.
Teach based on one-second-long sounds, from files or your microphone.
Teach based on images, from files or your webcam.


Teachable Machine
Class 1
Add Image Samples:
Webcam
Upload
5
6
Class 2
Add Image Samples:
Webcam
Upload
Training
Train Model
Advanced




شكل 4.5: إنشاء نموذج وتدريبه

يمكنك اختبار النموذج عند الانتهاء من عملية التدريب من خلال تقديم صورة من بيانات الاختبار، وذلك إما من الفئة الإيجابية (التي تحتوي على الشقوق) أو من الفئة السلبية (التي تخلو من الشقوق)، ثم يمكنك تقييم المخرجات.



## لاختبار وتقييم نموذج:

< حدِّد صـورة بهـا شـقوق في الخرسـانة مـن المجلد الفرعـي **اختبـار** (Test) الموجـود في مجلد **صور التصنيف** (Images for classification) داخل **المستندات** (Documents) وقُمْ بتحميلها. 1

< اضغـط علـى Choose images from your files, or drag & drop here (اختيـار الصـور مـن ملفاتـك أو اسـحبها وأفلتهـا هنـا). 2

شكل 4.6: اختبار وتقييم نموذج

كما تلاحظ، فقد صَنَّف النموذج الصورة بشكل صحيح في الفئة الإيجابية مع نسبة يقين 100 %، وذلك متوقّع؛ لأن صورة الخرسانة التي قمت بتحميلها تحتوي على الشقوق. وعليك تكرار الخطوتين الأخيرتين لتحميل صورة مختلفة وتقييم النموذج مرة أخرى.



## ميزات وتحديات النمذجة التنبؤية Benefits and Limitations of Predictive Modeling

**ميزات النمذجة التنبؤية:**

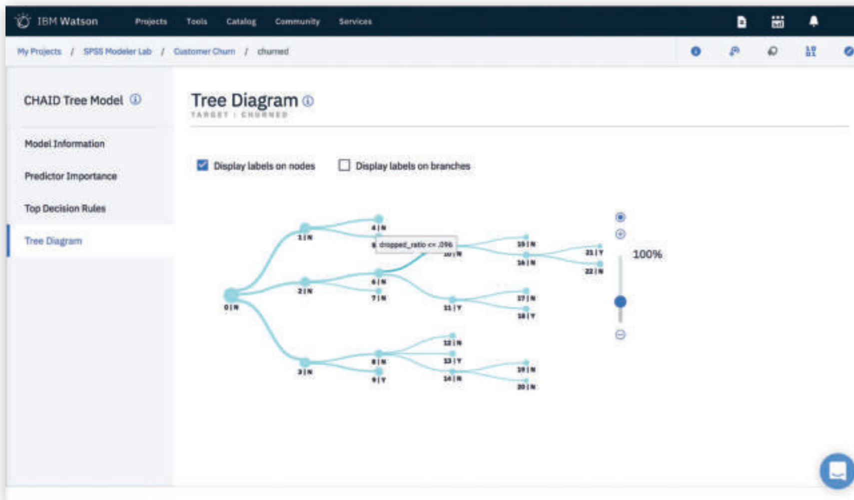
| |
|---|
| تحسين إستراتيجيات التسويق والمبيعات وخدمة العملاء. |
| تحسين التنافسية المبنية على المعرفة وتوظيف الإستراتيجيات لاكتساب ميزة المنافسة. |
| تعزيز جودة المنتجات والخدمات. |
| التحليل الدقيق لمتطلبات المستهلك. |
| توفير التوقُّعات للعوامل الخارجية التي تؤثر على الإنتاجية أو سير العمل. |
| المساهمة في إدارة المخاطر المالية والاستثمارية. |
| توفير التنبؤ بالمواد أو بالمخزون من المواد المختلفة. |
| التنبؤ بالتوجهات المستقبلية للأعمال. |
| دعم عملية إدارة القوى العاملة وتحليل المشاكل المتعلقة بها. |

**تحديات النمذجة التنبؤية:**

| |
|---|
| أمن وخصوصية البيانات. |
| التعامل مع حجم كبير من البيانات. |
| تحديات إدارة البيانات. |
| الحاجة المستمرة لتكييف النماذج مع القضايا والمشاكل المستجدة. |

## أدوات النمذجة التنبؤية
## Predictive Modeling tools

توجد أدوات النمذجة التنبؤية الحديثة على صورة مَنصات متكاملة تدعم تطوير الخوارزميات وتحليل البيانات وتقديم النتائج الموثوقة، ويَتمُّ استخدام هذه الأدوات من قِبَل الشركات والمؤسسات البحثية لإخراج استنتاجات دقيقة وشاملة يمكنها المساهمة في اتخاذ القرارات الفعالة.




شكل 4.7: مخطط تحليل البيانات وتحويلها

**الأدوات المتاحة:**

| |
|---|
| منصة H2O للذكاء الاصطناعي (H2O Driverless AI). |
| منصة IBM واتسون ستوديو (IBM Watson Studio). |
| منصة رابيد ماينر ستوديو (RapidMiner Studio). |
| منصة ساب للتحليلات السحابية (SAP Analytics Cloud). |
| منصة ساس (SAS). |
| منصة IBM الحزمة الإحصائية للعلوم الاجتماعية (IBM SPSS). |
| منصة أوراكل لعلم البيانات (Oracle Data Science). |



### جدول 4.3: تطبيقات النمذجة التنبؤية

| التطبيق | الوصف |
|---|---|
| المبيعات | يمكن أن يساهم التحليل التنبؤي في تحديد مكانة الشركة المالية من حيث المبيعات والأرباح، فمن خلال الكشف عن الحالات الشاذة والتباين في البيانات المالية السابقة للأقسام المختلفة في الشركة، يمكن للنمذجة تحديد الأقسام ذات الأداء المنخفض مثل قسم المبيعات، وهذا يؤدي إلى تحسين أداء الشركة وإدخال التحسينات على الأقسام أو العمليات بما يتناسب مع إستراتيجيات النمو والأداء المتميز. |
| التسويق | يمكن للشركات استهداف فئات معينة من العملاء بالحملات الترويجية لمنتجات أو خدمات معينة، وذلك من خلال التحليل والتنبؤ استنادًا إلى البيانات السابقة، كما يمكن لها أيضًا توقُّع استجابات ومتطلبات هؤلاء العملاء، وهنا يكمُن أحد الأسباب الرئيسة في قيام الشركات بجمع البيانات السابقة. تُعدُّ معرفة رغبات العملاء والتنبؤ بالمنتجات والخدمات التي يرغبون بالحصول عليها في المستقبل من أهم إستراتيجيات التسويق الحديثة. |
| وسائل التواصل الاجتماعي | تُعدُّ وسائل التواصل الاجتماعي مصدرًا أساسيًا للبيانات الضخمة غير المنظَّمة وغير المتجانسة، والتي تتكوَّن من مشاركة ملايين الأشخاص يوميًا في الحديث عن القضايا والمواضيع المختلفة، ويُعدُّ تحليل بياناتها من أكثر التطبيقات استخدامًا للنمذجَة التنبؤية، حيث يَسمحُ للمؤسسات والشركات باستكشاف اهتمامات العملاء وبالتالي تطوير خططها المستقبلية وفقًا لذلك. |
| تقييم المخاطر | تُستخدم النمذجة على نطاق واسع في المؤسسات المالية لتقييم المخاطر المتعلقة بتمويل الأفراد والأعمال، حيث تساهم بشكل فعال في تقييم أهليتهم للتمويل وفي الكشف عن الاحتيال، ويُمكن لأدوات التحليل التنبؤية أيضًا مساعدة المؤسسات في إجراء تقييم لمخاطر الاستثمار وتحديد درجة المخاطرة أو العائد المستقبلي على الاستثمار. |
| تحسين الجودة | تُستخدم النمذجة في عملية تحسين الجودة من خلال الاستعانة بملاحظات العملاء حول منتج أو خدمة معينة لتحسين جودتها، وكذلك للتنبؤ بالأثر المتوقَّع للتغييرات في المنتجات أو الخدمات من حيث زيادة المبيعات أو إقبال الزبائن على شرائها. |



وزارة التعليم
Ministry of Education
2022 - 1444

# تمرينات

**1**

| حدد الجملة الصحيحة والجملة الخاطئة فيما يلي: | صحيحة | خاطئة |
|---|---|---|
| 1. تَستخدم الشركات التحليلات التنبؤية للعثور على أنماط معينة في هذه البيانات تساعد في التعرف على المخاطر والفُرص المتاحة التي تتعلق بعمليات تلك الشركات. | ⚪ | ⚪ |
| 2. يزداد تعقيد النماذج ويصبح من الصعب تفسيرها بشكلٍ مبسط كلما ارتفعت دقة التحليلات المراد الحصول عليها. | ⚪ | ⚪ |
| 3. تُعدُّ البيانات ذات الطبيعة المعقدة، كتلك المتعلقة بالسلوك البشري أحد أسباب فشل النموذج. | ⚪ | ⚪ |
| 4. يُعدُّ الحصول على بيانات ذات علاقة بالنموذج التنبؤي من أول المتطلبات لنجاح ذلك النموذج وعمله بشكل فاعل. | ⚪ | ⚪ |
| 5. يُعدُّ تقييم المخاطر المالية من أهم استخدامات التحليل التنبؤي. | ⚪ | ⚪ |
| 6. لا يُمكن لنموذج التنبؤ التعامل مع أكثر من متغير واحد في نفس الوقت. | ⚪ | ⚪ |
| 7. تُستخدم نماذج القيم الشاذة لاكتشاف المُعامَلات الاحتيالية والحركات المشبوهة. | ⚪ | ⚪ |
| 8. يُمكن لنموذج السلاسل الزمنية تحليل العوامل الخارجية الموسمية أو العارضة التي يمكنها أن تؤثر على الاتجاهات المستقبلية. | ⚪ | ⚪ |
| 9. يُمكن وصف المُعامِل بأنه متغير جوهري في النموذج. | ⚪ | ⚪ |
| 10. تُستخدم نماذج التنبؤ لتقديم التوقُّعات المستقبلية لأحداث معينة من خلال استخدام قيم البيانات السابقة المتوفرة، وذلك ضِمن تسلسلٍ زمني محدد كعواملَ الإدخال في مجموعة البيانات. | ⚪ | ⚪ |



**2** اشرح بإيجاز مفهوم النمذجة التنبؤية مُستعينًا بالبحث على الإنترنت وأعطِ مثالًا عليها.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**3** اشرح بإيجاز خطوات البدء بإنشاء نموذج تنبؤي.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**4** صف بعض التطبيقات العملية للنمذجة التنبؤية في واقعنا اليومي.

______________________________________________

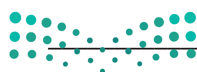
______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________



**5** تم تكليفك بإنشاء نموذج تنبؤي لحوادث المرور في المملكة، وبالطبع فإن عليك أولًا تحديد البيانات التي تحتاجها لهذا النموذج. ابحث في البوابة الوطنية للبيانات المفتوحة عبر الويب (https://data.gov.sa) عن البيانات المناسبة، ثم حدِّد نوع هذه البيانات وعدد سنوات البيانات التي ستحتاجها.

______________________________________________________________________________

______________________________________________________________________________

______________________________________________________________________________

______________________________________________________________________________

**6** تم تكليفك بإنشاء نموذج تنبؤي لتصنيف مجموعة من الصور لوسائل المواصلات المختلفة، والتي تشمل السيارات والطائرات والسفن. وضِّح الخطوات التي ستقوم بها لإنشاء هذا النموذج، بِدءاً من عملية جمع البيانات حتى عملية تدريب النموذج.

______________________________________________________________________________

______________________________________________________________________________

______________________________________________________________________________

______________________________________________________________________________

______________________________________________________________________________

**7** ابحث على الإنترنت عن أمثلة حول المخاوف الأخلاقية المتعلقة بالخصوصية، والتي قد تُعزَى إلى استخدام النمذجة التنبؤية. على سبيل المثال، هل يحق للشركات من الناحية الأخلاقية تعيين وترقية الموظفين وفقًا لنماذج التنبؤ التي تعتمد على البيانات الصحية لأولئك الموظفين؟

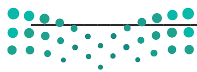
______________________________________________________________________________

______________________________________________________________________________

______________________________________________________________________________

______________________________________________________________________________

______________________________________________________________________________



رابط الدرس الرقمي



www.ien.edu.sa

الدرس الثاني

# التوقع (Forecasting)

التَوقُّع هو تقدير للأحداث المُستقبلية يتم إجراؤه بتضمين وتحليل البيانات السابقة بطريقة مَنهجية مُحددة مسبقًا. وتعتبر عملية تقدير المبيعات المحتملة أو الدخل المستقبلي أحد الأمثلة الشائعة على التوقُّع، حيث يتم استخدام بيانات المبيعات أو الدَخل السابق كمرجع لتَوقُّع المبيعات المستقبلية. ويُمكن تطبيق التوقُّع في الكثير من المجالات الأُخرى في الحياة، كتوقُّع مقدار النمو السكاني للأعوام القادمة، أو توقُّع عدد السياح الذين سيزورون المملكة العربية السعودية خلال فترة محددة، وما إلى ذلك.

**التوقُّع (Forecasting):**

هو وضع تقديرات لأحداث مستقبلية بناءً على معلومات أو بيانات سابقة.

رغم وجود بعض أوجه التشابه بين مصطلحي التوقُّع والتنبؤ، إلا إنهما غير متطابقين.

إن التنبؤ هو عملية إنشاء نموذج لتخمين أو تقدير نتيجة أو حدث ما بناءً على قيم مُتغيرات حالية، بينما التوقُّع هو عملية تقدير لقيمة متغير ما في المستقبل بناءً على القيم السابقة لنفس المتغير في إطار زمني مُحدد. يَعني هذا أن التوقُّع هو نوع من أنواع التنبؤ يرتبط بالزمن. فالتوقُّع يشير إلى السلاسل الزمنية والمستقبلية، بينما يُعنى التنبؤ بعوامل أخرى غير الزمن. وبمعنىً آخر، فإنك عندما تتنبأ بالمستقبل فأنت تتوقَّع، ويمكنك القول بأن جميع التوقُّعات هي في الواقع عبارة عن تنبؤات، بينما لا يمكن وصف جميع التنبؤات بأنها توقُّعات، كتلك الحالات التي يتم فيها استخدام الانحدار لشرح العلاقة بين متغيرين على سبيل المثال.

وهكذا فقد أصبحتَ تُدركُ خطأ تسمية عملية توقُّع الطَقس بالتنبؤ بالطَقس، فتوقُّع الطقس يتنبأ بالطقس في المستقبل باستخدام معلومات زمنية محددة. على سبيل المثال، إذا كانت السماء تمطر الآن، فما هو احتمال استمرار هطول الأمطار في غُضون الدقائق الخمس القادمة؟ من الممكن القول بثقة أنه وبصَرف النظر عن جميع العوامل الأخرى التي تؤثر على الطقس (مثل الضغط الجوي ودرجة الحرارة)، فإنَ احتمالية استمرار هطول الأمطار في غضون خمس دقائق ستكون مُرتفعة، لأنها تمطر في الوقت الحالي.

تتيح تقنيات التوقُّع للشركات والمؤسسات تعديل سياساتها واتخاذ القرارات السليمة لتحقيق أهدافها.

**يجب اتباع الخطوات التالية في عملية التوقُّع:**

**جدول 4.4: خطوات التوقُّع**

| | وصف الخطوة |
|---|---|
| خطوة 1 | تحديد البيانات المراد تحليلها والسعي للحصول عليها. |
| خطوة 2 | استخدام أدوات برمجية لإعداد مجموعة البيانات. |
| خطوة 3 | ضَبط السلسلة الزمنية المناسبة للتوقُّع. |
| خطوة 4 | إنشاء التوقُّعات. |
| خطوة 5 | تَمثيل البيانات رسوميًا. |
| خطوة 6 | تَحليل النتائج. |



## التوقُّع في برنامج جداول البيانات إكسل Forecasting in Excel

هناك العديد من الأدوات البرمجية التي يُمكن استخدامها لإنشاء نموذج تحليلي للبيانات السابقة للقيام بعملية التوقُّع، ويُعتبر **مايكروسوفت إكسل** (Microsoft Excel) من أهم هذه الأدوات، ويُمكن من خلاله استخدام البيانات السابقة المُستندة إلى الزمن من أجل إنشاء توقُّع. ستتعلم في هذا الدرس خطوات تَوقُّع أعداد الرحلات السياحية المستقبلية إلى المملكة العربية السعودية وذلك باستخدام برنامج إكسل، حيث تقوم تحديدًا بإجراء التوقُّع في هذا البرنامج للتنبؤ بأعداد الرحلات السياحية المتوقّعة لكل شهر من شهور العام 2023، بناءً على المعطيات السابقة للبيانات السياحية للعام 2019.

### خطوة 1: الحصول على البيانات Obtain the data

يتعين عليك في البداية الحصول على البيانات المطلوبة لتحليل التوقُّعات؛ لأن الهدف هو توقُّع العدد الشهري للرحلات السياحية للعام 2023، ولذلك أنت بحاجة إلى البيانات السابقة لأعداد تلك الرحلات، ولهذا الغرض، سَتجمَع هذه البيانات من صفحة بيانات السياحة التابعة لوزارة السياحة السعودية من خلال البوابة الوطنية للبيانات المفتوحة (https://data.gov.sa). يُمكنك الحصول على البيانات الشهرية السياحية للعام 2019 من خلال الرابط التالي:

https://data.gov.sa/Data/ar/dataset/the-main-indicators-of-tourism-demand-statstics-2015-2020.

من المهم أن تلاحظ أنه لا يمكن استخدام بيانات العامين 2020 و2021 بسبب تأثيرات جائحة **كورونا** (COVID-19) على السياحة والسفر، ولهذا السبب ستلجأ لإجراء حسابات التوقُّع باستخدام بيانات العام 2019.

امسح رَمز الاستجابة السريعة (QR code) لتحميل ملف البيانات.



هذه هي بيانات السياحة للمملكة العربية السعودية لعام 2019 المُقدمة من صفحة بيانات السياحة (وزارة السياحة).

شكل 4.8: البوابة الوطنية للبيانات المفتوحة

## لتحميل البيانات:

- اضغط على زر **تحميل** (Download). 1
- في نافذة **التنزيلات** (Downloads)، اضغط على **فتح ملف** (Open File). 2
- في ورقة العمل I1، حدِّد وانسخ محتويات الخلايا J59:J70، 3 وألصقها في الخلايا A2: A13 في ملف إكسل جديد. 4
- اكتُب "الشهر" في الخلية A1 وأضف "2019" إلى اسم كل شهر في الخلية. 5
- في ورقة العمل I1 من ملف إكسل الذي تم تحميله، حدد وانسخ محتويات الخلايا D59:D70، 6 وألصقها في الخلايا B2: B13 في ملف الإكسل. 7
- اكتُب "الرحلات السياحية" في الخلية B1. 8



حفظ تلقائي | tic_tourist-demand-indicators-2015-2020 - طريقة عرض محمية | Binary Academy

ملف | الشريط الرئيسي | إدراج | تخطيط الصفحة | الصيغ | بيانات | مراجعة | عرض | تعليمات | التعليقات | مشاركة

طريقة عرض محمية كن حذراً—قد تحتوي الملفات من إنترنت على فيروسات. يعتبر البقاء في "طريقة عرض محمية" أكثر أماناً إلا إذا كنت تريد تحريرها. تمكين التحرير

D59 | fx | 1653.34332432955

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 43 | | | November | 1,833 | 17,679 | 10,445 | 9.6 | 5,699 | 591 | نوفمبر | |
| 44 | | | December | 1,381 | 11,586 | 7,228 | 8.4 | 5,234 | 624 | ديسمبر | |
| 45 | | | **Total** | **16,109** | **171,036** | **97,778** | **10.6** | **6,070** | **572** | الإجمالي | |
| 46 | | | January | 1,391 | 11,257 | 6,113 | 8.1 | 4,395 | 543 | يناير | |
| 47 | | | February | 1,182 | 9,667 | 6,000 | 8.2 | 5,076 | 621 | فبراير | |
| 48 | | | March | 1,496 | 17,043 | 6,865 | 11.4 | 4,588 | 403 | مارس | |
| 49 | | | April | 1,511 | 14,069 | 5,688 | 9.3 | 3,764 | 404 | أبريل | |
| 50 | | | May | 1,619 | 16,287 | 6,732 | 10.1 | 4,157 | 413 | مايو | |
| 51 | | 2018 | June | 1,070 | 9,302 | 4,462 | 8.7 | 4,172 | 480 | يونيو | 2018 |
| 52 | | | July | 641 | 3,332 | 2,657 | 5.2 | 4,146 | 797 | يوليو | |
| 53 | | | August | 2,444 | 49,562 | 30,479 | 20.3 | 12,473 | 615 | أغسطس | |
| 54 | | | September | 509 | 2,720 | 2,303 | 5.3 | 4,521 | 847 | سبتمبر | |
| 55 | | | October | 903 | 8,850 | 5,492 | 9.8 | 6,082 | 621 | أكتوبر | |
| 56 | | | November | 1,141 | 14,707 | 7,933 | 12.9 | 6,954 | 539 | نوفمبر | |
| 57 | | | December | 1,427 | 17,132 | 8,753 | 12.0 | 6,134 | 511 | ديسمبر | |
| 58 | | | **Total** | **15,334** | **173,929** | **93,478** | **11.3** | **6,096** | **537** | الإجمالي | |
| 59 | | | January | 1,653 | 14,339 | 7,982 | 8.7 | 4,828 | 557 | يناير | |
| 60 | | | February | 1,492 | 13,385 | 6,775 | 9.0 | 4,541 | 506 | فبراير | |
| 61 | | | March | 1,599 | 16,688 | 6,762 | 10.4 | 4,228 | 405 | مارس | |
| 62 | | | April | 1,650 | 15,236 | 6,211 | 9.2 | 3,763 | 408 | أبريل | |
| 63 | | | May | 2,079 | 19,757 | 10,167 | 9.5 | 4,889 | 515 | مايو | |
| 64 | | 2019 | June | 903 | 6,898 | 4,056 | 7.6 | 4,489 | 588 | يونيو | 2019 |
| 65 | | | July | 750 | [illegible] | 3,363 | 6.5 | 4,485 | 693 | يوليو | |
| 66 | | | August | 2,675 | [illegible] | 34,911 | 19.6 | 13,053 | 667 | أغسطس | |
| 67 | | | September | 635 | [illegible] | 3,145 | 6.5 | 4,952 | 767 | سبتمبر | |
| 68 | | | October | 1,119 | 11,081 | 5,199 | 9.9 | 4,647 | 469 | أكتوبر | |
| 69 | | | November | 1,234 | 13,247 | 5,334 | 10.7 | 4,322 | 403 | نوفمبر | |
| 70 | | | December | 1,735 | 17,103 | 9,448 | 9.9 | 5,445 | 552 | ديسمبر | |
| 71 | | | **Total** | **17,526** | **189,036** | **103,354** | **10.8** | **5,897** | **547** | الإجمالي | |
| 72 | | | January | 1,737 | 16,211 | 8,806 | 9.3 | 5,070 | 543 | يناير | |
| 73 | | | February | 1,483 | 15,528 | 8,010 | 10.5 | 5,402 | 516 | فبراير | |
| 74 | | | March | 192 | 1,296 | 796 | 6.8 | 4,153 | 614 | مارس | |
| 75 | | | April | 27 | 71 | 64 | 2.6 | 2,354 | 902 | أبريل | |
| 76 | | | May | 19 | 49 | 48 | 2.6 | 2,543 | 969 | مايو | |

Content | G1 | G2 | I1 | I2 | I3 | I4 | I5 | I6 | I7 | I8 | D1 | D2 | D3 | D4 | D5 | D6 | D7 | O1 | O2 | O3 | O4 | O5 | O6 | Defir ...

80% | مجموع: 17,526 | عدد: 24 | متوسط: 1,460 | جاهز

3

6

A1 | fx

| | A | B |
|---|---|---|
| 1 | الشهر | الرحلات السياحية |
| 2 | يناير 2019 | 1,653 |
| 3 | فبراير 2019 | 1,492 |
| 4 | مارس 2019 | 1,599 |
| 5 | أبريل 2019 | 1,650 |
| 6 | مايو 2019 | 2,079 |
| 7 | يونيو 2019 | 903 |
| 8 | يوليو 2019 | 750 |
| 9 | أغسطس 2019 | 2,675 |
| 10 | سبتمبر 2019 | 635 |
| 11 | أكتوبر 2019 | 1,119 |
| 12 | نوفمبر 2019 | 1,234 |
| 13 | ديسمبر 2019 | 1,735 |
| 14 | | |

2019

120% | جاهز | إمكانية الوصول لذوي الاحتياجات الخاصة: جاهز لما تريده

5

8

7

4

شكل 4.9: تحميل البيانات

## خطوة 2: استخدام أداة التوقُّع Use a forecasting tool

بعدَ الحُصول على بيانات أعداد الرحلات السياحية الشهرية لعام 2019، يَتعيَّن عليك تصديرها إلى أداة برمجية للتوقُّع. ستَستخدم مايكروسوفت إكسل لهذا الغرض، وستُنشئُ عمودين في ورقة جديدة تسمى "2019". سيحتوي العمود الأول على أسماء أشهر السنة الميلادية، والعمود الثاني على بيانات أعداد الرحلات السياحية لكل شهر من أشهر العام 2019.

**السلاسل الزمنية (Time series):**

هي مجموعة من البيانات الكمية، وتمثل وصفًا لواحدة أو أكثر من الصفات أو الخصائص المتعلقة بشخص أو ظاهرة أو حدث معين، والتي يتم جمعها خلال فترات زمنية متتابعة.

## خطوة 3: ضبط السلاسل الزمنية Set the time series

تتطلبُ عملية إنشاء التوقُّع في مايكروسوفت إكسل وجود عمود يحتوي على سلسلة زمنية بقِيَم رقمية (أرقام أو تواريخ)، ولهذا السبب لا يمكنك استخدام العمود الذي يحتوي علىً أسماء الشهور كعمود سلسلة زمنية وذلك لِكَون البيانات فيه نصيَّة. وللتغلُّب على هذه المشكلة، ستُنشئُ عمودًا يُمَثل الشهور ويحتوي على الأرقام من 1 إلى 12 بشكلٍ متسلسل كما هو موضح في شكل 4.10.

| | الشهر | الرحلات السياحية |
|---|---|---|
| 1 | يناير 2019 | 1,653 |
| 2 | فبراير 2019 | 1,492 |
| 3 | مارس 2019 | 1,599 |
| 4 | أبريل 2019 | 1,650 |
| 5 | مايو 2019 | 2,079 |
| 6 | يونيو 2019 | 903 |
| 7 | يوليو 2019 | 750 |
| 8 | أغسطس 2019 | 2,675 |
| 9 | سبتمبر 2019 | 635 |
| 10 | أكتوبر 2019 | 1,119 |
| 11 | نوفمبر 2019 | 1,234 |
| 12 | ديسمبر 2019 | 1,735 |



شكل 4.10: البيانات في ورقة عمل إكسل

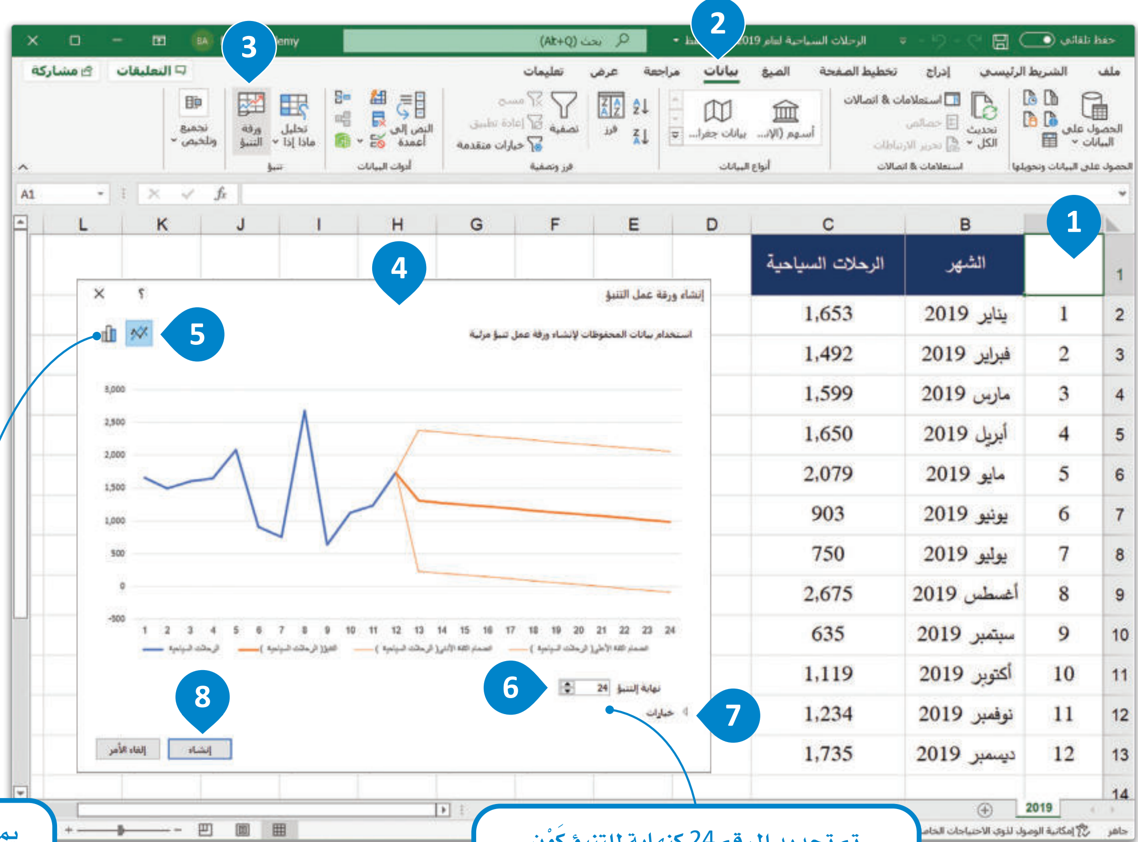
## خطوة 4: إنشاء التوقُّع Create the forecast

استـنادًا إلى بيانـات أعـداد الرحـلات السـياحية الشـهرية لعـام 2019، ستَسـتخدم خيـار **ورقة التنبـؤ** (Forecast sheet) مـن علامـة تبويب **بيانات** (Data) في مايكروسـوفت إكسـل مـن أجل إنشـاء التوقُّـع.

### لإنشاء التوقُّع:

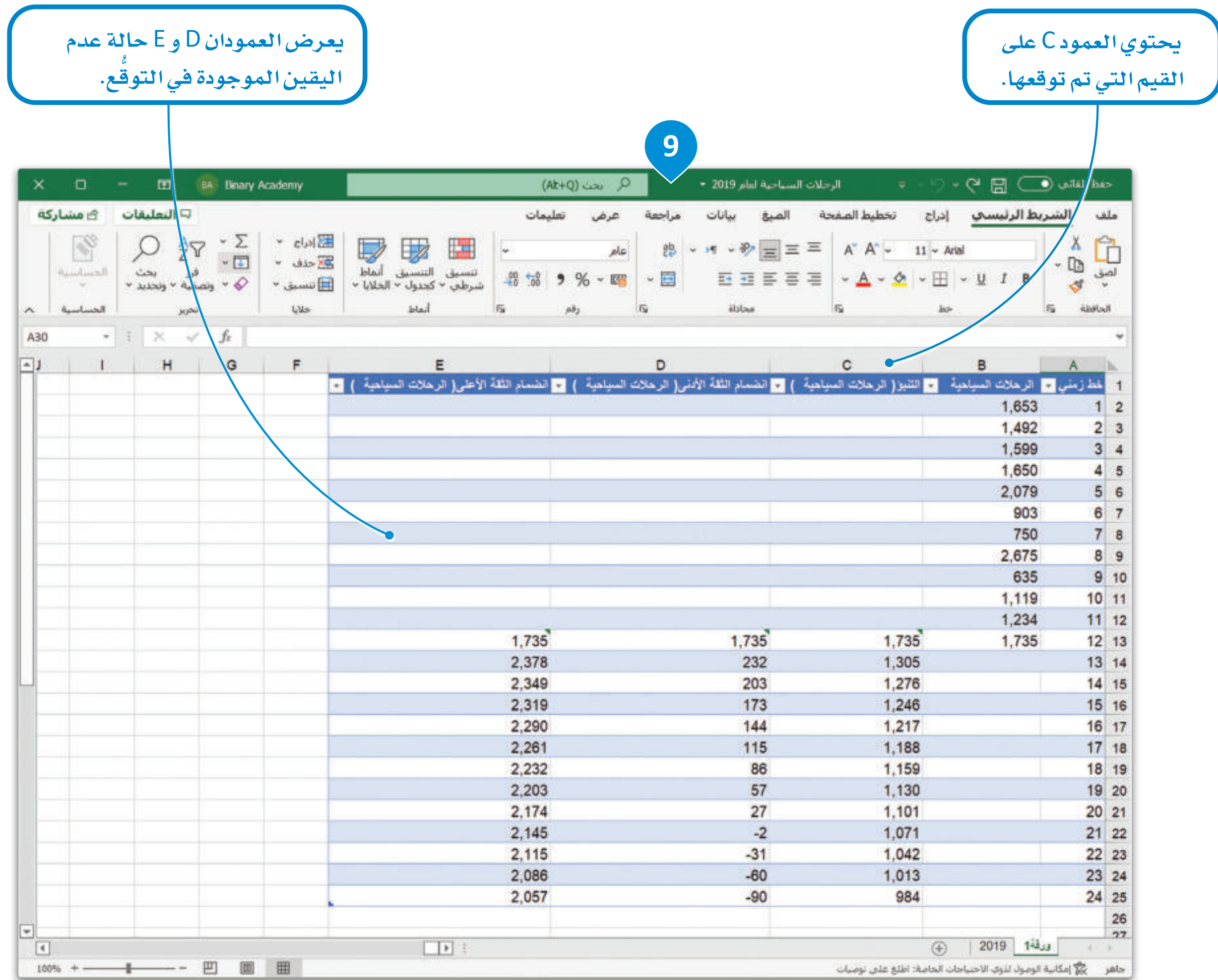
- اضغط على الخلية A1. 1
- اضغط على علامـة تبويب **بيانات** (Data)، 2 ومن مجموعـة **تنبؤ** (forecast)، اضغط على **ورقة التنبـؤ** (Forecast sheet). 3
- ستظهر نافذة **إنشاء ورقة عمل التنبؤ** (Create Forecast Worksheet). 4
- اختر **المخطط الخطي** (Line chart). 5
- اضبط قيمة **نهاية التنبؤ** (Forecast end) على 24. 6
- اضغط على **خيارات** (Options)، 7 لعمل التغييرات في إعدادات التوقُّع الإضافية.
- اضغط على **إنشاء** (Create). 8
- ستجد ورقة العمل الجديدة التي تحتوي على قِيم التوقُّع إلى يمين الورقة حيث أدخلت سلسلة البيانات.

**يستخدم مايكروسوفت إكسل مصطلح مختلف لمصطلح التوقُّع (Forecast) وهو التنبؤ.**



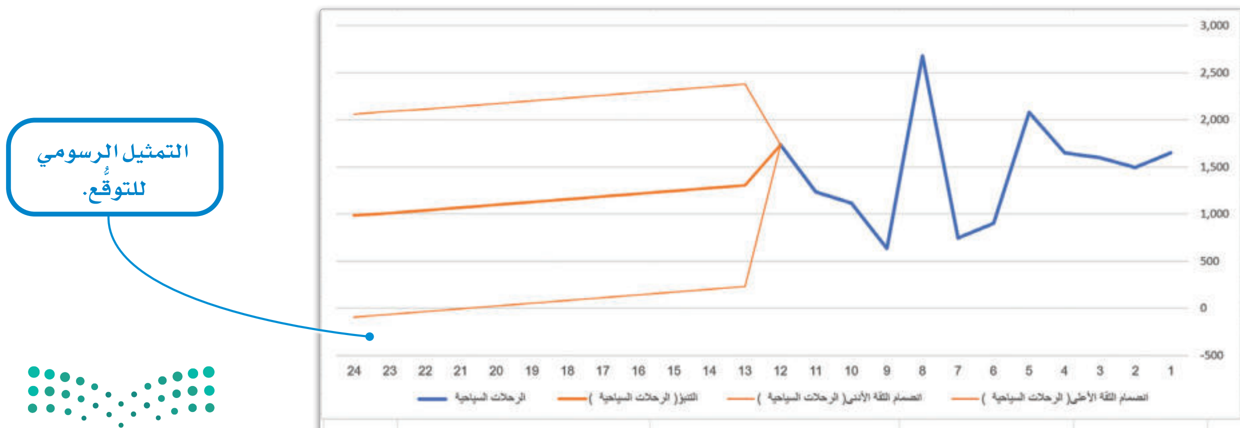

يمكن عرض التوقُّع باستخدام مخطط خطي أو مخطط عمودي.

تم تحديد الرقم 24 كنهاية للتنبؤ كَوْن البيانات المتوفرة هي لمدة 12 شهرًا، وتريد توقُّعًا للأشهر 12 القادمة، وبالتالي نتجت القيمة من: 12 + 12 = 24.




| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | خط زمني | الرحلات السياحية | التنبؤ( الرحلات السياحية ) | انضمام الثقة الأدنى( الرحلات السياحية ) | انضمام الثقة الأعلى( الرحلات السياحية ) |
| 2 | 1 | 1,653 | | | |
| 3 | 2 | 1,492 | | | |
| 4 | 3 | 1,599 | | | |
| 5 | 4 | 1,650 | | | |
| 6 | 5 | 2,079 | | | |
| 7 | 6 | 903 | | | |
| 8 | 7 | 750 | | | |
| 9 | 8 | 2,675 | | | |
| 10 | 9 | 635 | | | |
| 11 | 10 | 1,119 | | | |
| 12 | 11 | 1,234 | | | |
| 13 | 12 | 1,735 | 1,735 | 1,735 | 1,735 |
| 14 | 13 | | 1,305 | 232 | 2,378 |
| 15 | 14 | | 1,276 | 203 | 2,349 |
| 16 | 15 | | 1,246 | 173 | 2,319 |
| 17 | 16 | | 1,217 | 144 | 2,290 |
| 18 | 17 | | 1,188 | 115 | 2,261 |
| 19 | 18 | | 1,159 | 86 | 2,232 |
| 20 | 19 | | 1,130 | 57 | 2,203 |
| 21 | 20 | | 1,101 | 27 | 2,174 |
| 22 | 21 | | 1,071 | -2 | 2,145 |
| 23 | 22 | | 1,042 | -31 | 2,115 |
| 24 | 23 | | 1,013 | -60 | 2,086 |
| 25 | 24 | | 984 | -90 | 2,057 |

عندما تُنشئ توقُّعًا، يُنشئ إكسل ورقة عمل جديدة تحتوي على جدول قيم البيانات السابقة وقيم البيانات (المستقبلية) التي تم توقعها، وكذلك يتم عرض مقدار **عدم اليقين** (Uncertainity) أيضًا مع **الحد الأعلى للثقة** (Upper Confidence Bound) و**الحد الأدنى للثقة** (lower Confidence Bound)، ويقوم إكسل بإنشاء المخطط المحدد لتمثيل هذه البيانات.

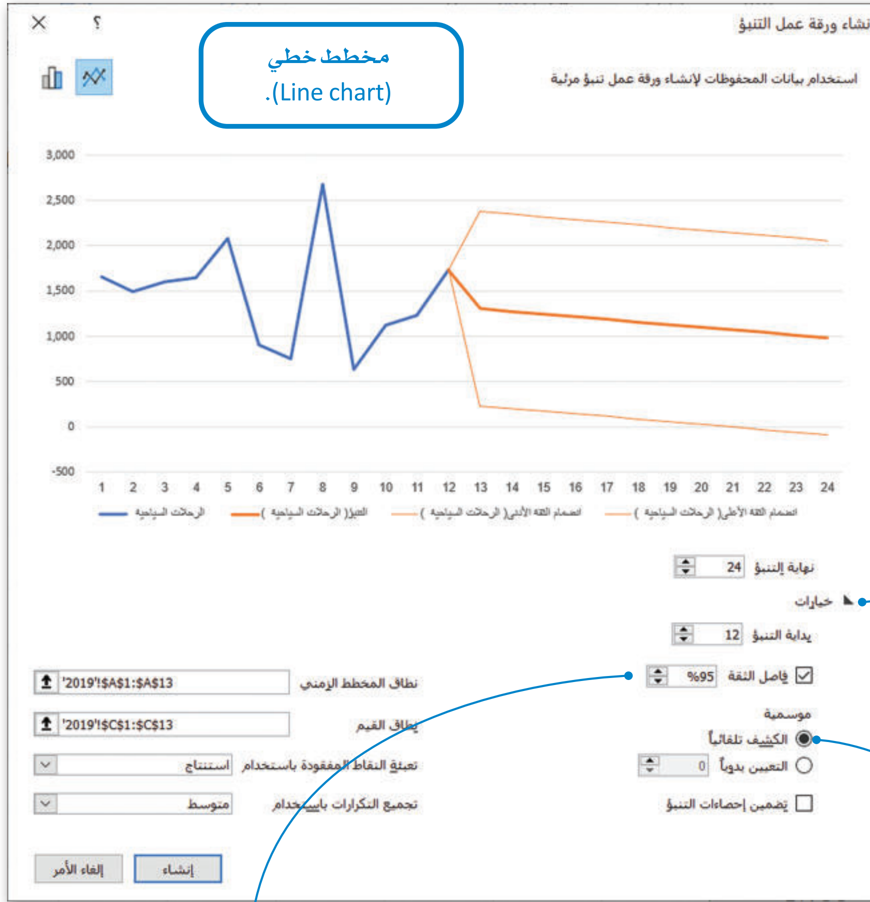



شكل 4.11: إنشاء التوقُّع



## إعدادات التوقُّع الإضافية Additional forecast settings

يَستخدم النموذج برنامج إكسل لتوقُّع القيم المستقبلية للبيانات بناءً على القيم الموجودة (البيانات السابقة) باستخدام **الانحدار الخطي** (Linear Regression) وهو نموذج إحصائي لكشف العلاقة بين متغيرين مستمرين، حيث يتم التنبؤ من خلاله بقيم مُتغير تابع بناءً على قيم مُتغير مستقل (في المثال السابق المتعلق بالسياحة يمثل عدد الرحلات السياحية المتغير التابع، بينما يكون المتغير المستقل هو الوقت ممثلًا بالشهور). ويُعدُّ الانحدار الخطي نوعًا أساسيًا وشائع الاستخدام في التحليل التنبؤي والذي يتيح فهم ودراسة العلاقة الكمية بين متغيرين مستمرين.




إذا قمت بالضغط على خيارات (Options)، فستظهر قائمة منسدلة تُمكّنك من تغيير الإعدادات المتقدمة الخاصة بالتنبؤ.

شكل 4.12: إعدادات التوقُّع الإضافية

يحدد فاصل الثقة (Confidence interval) دقة التنبؤ.

من خلال اختيار الكشف تلقائيًا (Detect Automatically)، يُمكن لبرنامج إكسل تعيين قيمة الموسمية (Seasonality) تلقائيًا إلى 12.

على الرغم من أن الانحدار الخطي هو الأكثر شيوعًا وموثوقية للتنبؤ في النموذج، إلا أنه يَفتقر إلى إمكانية التعامل مع البيانات النوعية. فعلى سبيل المثال، قد تَمثل بعض البيانات النوعية في المثال المتعلق بالسياحة في تفضيلات السائحين بخصوص توقيت حصولهم على العطلة أو الإجازة، وتُعدُّ هذه البيانات ذات تأثير في مسألة اختيار زمان ومكان قضاء العطلة. قد تعمل دالة التوقُّع القائمة على الانحدار الخطي في بعض الأحيان، ولكن عدم تعامل الانحدار الخطي مع البيانات النوعية يجعل الكثير من التوقُّعات بعيدة عن الواقع بشكل كبير، مما يؤثر سلبًا على التنبؤات بشكل عام.

يستخدم مايكروسوفت إكسل مصطلح مختلف لمصطلح نطاق الثقة (Confidence interval) وهو فاصل الثقة.

## نطاق الثقة Confidence Interval

يَنطوي أي توقُّع باستخدام النمذجة على قَدْرٍ مُعيّنٍ من الشك وعدم التيقن بحدوثه، فالتوقُّع كما تعلمت لا يتضمن قيمًا حقيقية تم جمعها أو قياسها من خلال البحث مثلاً، بل هي قيم تقديرية غير موجودة بالفعل بعد.

تحتمل القيم التي يتم تخمينها الصواب أو الخطأ في المستقبل بغضّ النظر عن الطريقة التي تم تخمينها بها، ويعطي نطاق الثقة مجموعة من القيم المتوقّعة وليس فقط قيمة متوقّعة واحدة، و يتم تحديد هذه الفترة من خلال الحد الأدنى والحد الأعلى للثقة، وهكذا فإن القيمة الفعلية يجب ألا تقل عن قيمة الحد الأدنى للثقة وألا تزيد عن الحد الأعلى للثقة.

تُعرَف هذه القيم إحصائيًا باسم نطاق الثقة وتمثِّل نطاقًا من القيم المُقدَّرة لمتغير ما، ويمكن النظر إليها على أنها متوسط القيمة التي يتم تخمينها للمتغير ± نطاق الاختلاف في هذا التخمين.

يتم حساب نطاق الثقة وفق مستوىً محدد، وعادة ما يساوي 95%، ويعني هذا أن القيمة الحقيقية لديها فرصة بنسبة 95% لتقع بين نطاق القيم المتوقّعة بين الحد الأدنى للثقة والحد الأعلى للثقة.

**نطاق الثقة (Confidence Interval):**

لنطاق الثقة احتمالٌ معروف ومحكوم عمومًا بنسبة تتراوح بين 95% و 99% من القيمة الحقيقية.

يشير التنبؤ الذي قُمت به في إكسل بخصوص أعداد الرحلات السياحية إلى أنه من المتوقّع وصول 1,305,000 رحلة سياحية في شهر يناير 2023، مع حد أدنى للثقة مقداره 232,000، وحد أعلى للثقة مقداره 2,378,000. وهكذا فإن نطاق الثقة ينحصر في القيم الواقعة ما بين 232,000 و 2,378,000. ويُحدَّد مستوى الثقة في إجراء التوقُّع في إكسل بنسبة 95%، وبناء على ذلك فإن عدد الرحلات السياحية المتوقّعة لشهر يناير 2023 تحظى بفرصة بنسبة 95% لأن تكون بين 232،000 و 2,378,000.

إذا كان عدد الرحلات السياحية في شهر يناير 2023 يساوي 1000،000، فهذا يعني أن التنبؤ كان صحيحًا تمامًا. ورغم أن القيمة 1000،000 لا تساوي القيمة 1,305,000 التي حصلنا عليها من التوقُّع، إلا أن تلك القيمة تقع في نطاق القيمتين 232,000 و 2,378,000، أي ضمن نطاق الثقة.

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | خط زمني | الرحلات السياحية | التنبؤ (الرحلات السياحية) | انضمام الثقة الأدنى (الرحلات السياحية) | انضمام الثقة الأعلى (الرحلات السياحية ) |
| 2 | 1 | 1,653 | | | |
| 3 | 2 | 1,492 | | | |
| 4 | 3 | 1,599 | | | |
| 5 | 4 | 1,650 | | | |
| 6 | 5 | 2,079 | | | |
| 7 | 6 | 903 | | | |
| 8 | 7 | 750 | | | |
| 9 | 8 | 2,675 | | | |
| 10 | 9 | 635 | | | |
| 11 | 10 | 1,119 | | | |
| 12 | 11 | 1,234 | | | |
| 13 | 12 | 1,735 | 1,735 | 1,735 | 1,735 |
| 14 | 13 | | 1,305 | 232 | 2,378 |
| 15 | 14 | | 1,276 | 203 | 2,349 |
| 16 | 15 | | 1,246 | 173 | 2,319 |
| 17 | 16 | | 1,217 | 144 | 2,290 |
| 18 | 17 | | 1,188 | 115 | 2,261 |
| 19 | 18 | | 1,159 | 86 | 2,232 |
| 20 | 19 | | 1,130 | 57 | 2,203 |
| 21 | 20 | | 1,101 | 27 | 2,174 |
| 22 | 21 | | 1,071 | -2 | 2,145 |
| 23 | 22 | | 1,042 | -31 | 2,115 |
| 24 | 23 | | 1,013 | -60 | 2,086 |
| 25 | 24 | | 984 | -90 | 2,057 |
| 26 | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| 13 | 12 | 1,735 | 1,735 | 1,735 | 1,735 |
| 14 | 13 | | 1,305 | 232 | 2,378 |
| 15 | 14 | | 1,276 | 203 | 2,349 |

شكل 4.13: نطاق الثقة



## خطوة 5: تمثيل البيانات رسوميًا Graph the data

كما تعلمت سابقًا، يُمكن تمثيل التوقُّعات باستخدام **مخطط خطي** (Line Chart) أو **مخطط عمودي** (Column Chart):

### المخطط الخطي Line Chart

تُستخدم المخططات الخطية على نطاقٍ واسعٍ لتمثيل التغيُّرات التي حدثت على قيم مُعَينة بمرور الوقت، وذلك على صورة سِلسلة من نقاط البيانات المتصلة معًا بخطوط مستقيمةٍ، وتُساعِد في تحديد العلاقة بين مجمُوعتين من القيم، كمجموعة شهور العام التي تمثل قيم الوقت، ومجموعة عدد الرحلات السياحية، وبالطبع لا بد من أن تكون مجموعة بيانات تعتمد دائمًا على المجموعة الأخرى (في المثال السابق يعتمد عدد الرحلات السياحية على الوقت).

| فوائد المخططات الخطية: |
| --- |
| تسمح بالتحليل السريع للبيانات. |
| تسمح بمراقبة التغييرات بسهولة خلال فترة زمنية معينة. |
| يمكن الاعتماد عليها لتمثيل مجموعات البيانات التي تحتوي على 50 قيمة كحد أعلى. |
| تُساعد في استنباط تنبؤات حول نتائج بيانات لم يتم تسجيلها بعد. |

### المخطط العمودي Column Chart

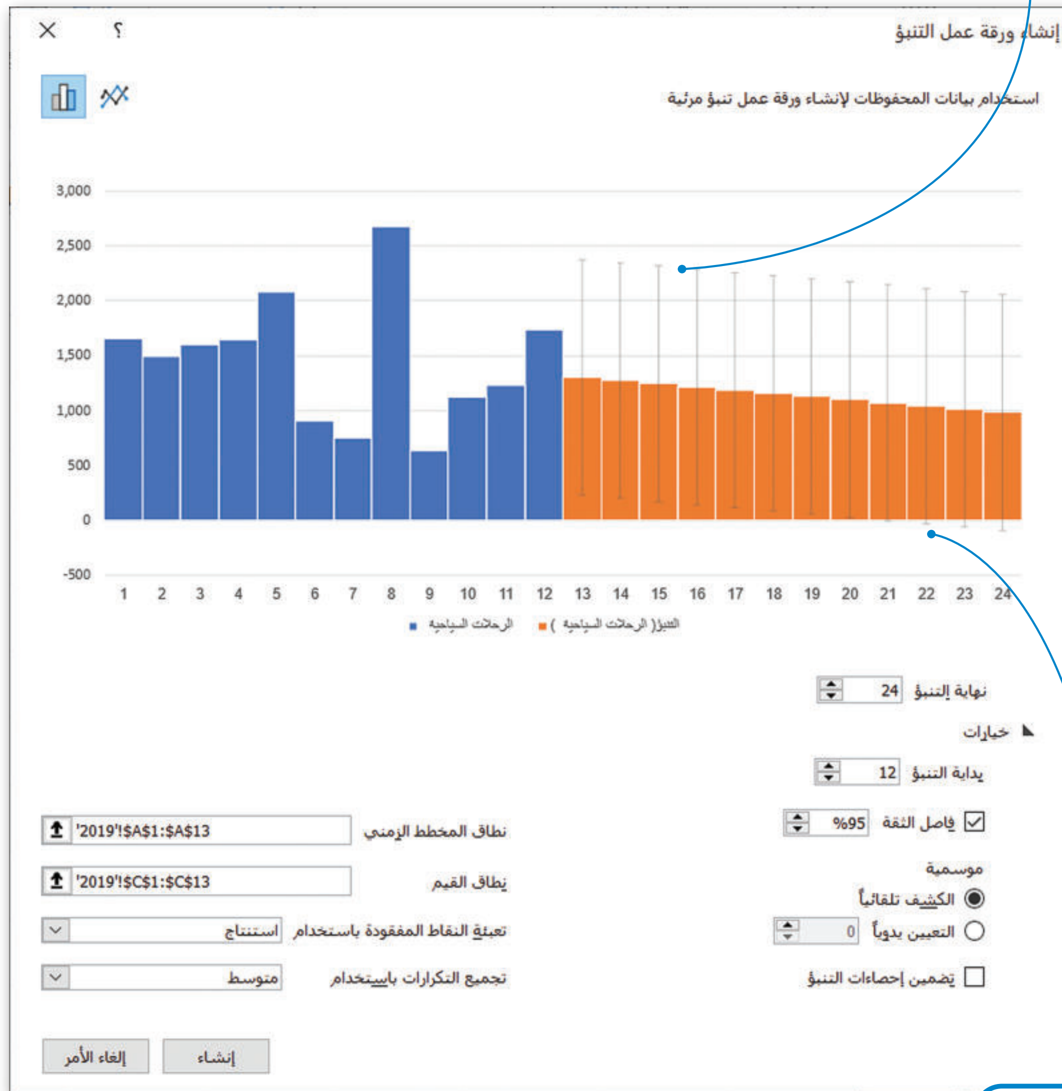
تُستخدم المخططات العمودية لعرض البيانات التي تم جمعها من خلال الاستبانات والمقابلات مثل بيانات الفئات العمرية وأعداد المنتجات المفضلة للزبائن وغيرها، ويُمكن استخدامها أيضًا مع البيانات مثل مجموعات البيانات الأخرى كالبيانات الشهرية لعدد الرحلات السياحية، شَرطَ ألا تحتوي هذه المجموعات على عددٍ كبيرٍ من القيم.

| فوائد المخططات العمودية: |
| --- |
| تتيح المقارنة بين مجموعات البيانات بسهولة. |
| يُمكن من خلالها تلخيص كمية كبيرة من البيانات بصورة مرئية يسهل تفسيرها. |
| تُسهِّل من ملاحظة التوجهات. |
| تُساعد على دراسة الأنماط في البيانات على امتداد فترة طويلة من الزمن. |



**تمنحك ورقة عمل التنبؤ في إكسل الاختيار بين المخطط الخطي والمخطط العمودي لتمثيل البيانات، ولقد تم استخدام المخطط الخطي في المثال المتعلق بالسياحة، وذلك لكونه أكثر ملاءمة للتمثيل المرئي للبيانات.**

تُشير الحواف العلوية لجميع الخطوط الرمادية إلى قيم الحد الأعلى للثقة.




شكل 4.14: المخطط العمودي

تُشير الحواف السفلية لجميع الخطوط الرمادية إلى قيم الحد الأدنى للثقة.



## تخصيص الرسومات Customize the Graphics

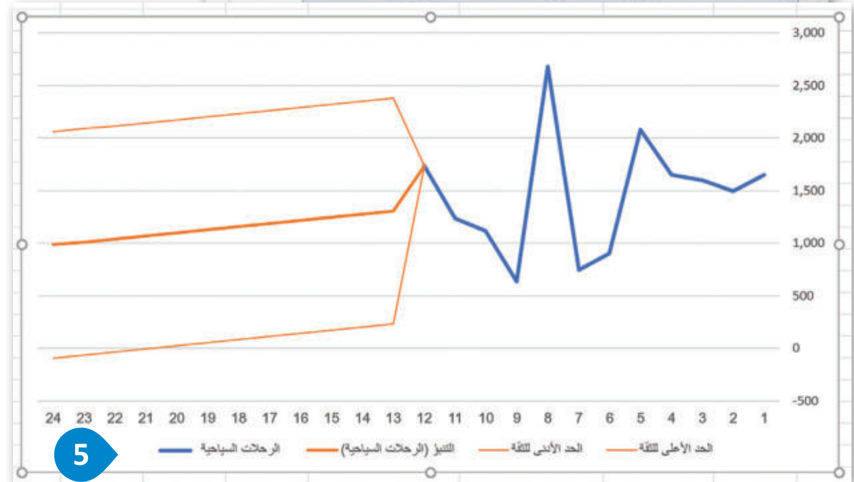
يقوم برنامج إكسل بإنشاء ورقة العمل الجديدة بقيم التوقُّع وأعمدة للحد الأعلى للثقة و الحد الأدنى للثقة حسب نموذج ثابت، ولكن يُمكنك تحرير أسماء الأعمدة وتعديلها.

### لتغيير أسماء الأعمدة:

- في **ورقة عمل التنبؤ**، اضغط على الخلية D1. **1**
- حدِّد الكلمات الموجودة في الخلية، واكتب "الحد الأدنى للثقة" واضغط على **Enter ↵**. **2**
- اضغط على الخلية E1، واكتب "الحد الأعلى للثقة" واضغط على **Enter ↵**. **3**
- سيتم تطبيق التغييرات في **ورقة عمل التنبؤ**، **4** وعلى **المخطط الخطي** أيضًا. **5**

| A خط زمني | B الرحلات السياحية | C التنبؤ (الرحلات السياحية) | D انضمام الثقة الأدنى (الرحلات السياحية) | E انضمام الثقة الأعلى (الرحلات السياحية) |
|---|---|---|---|---|
| 1 | 1,653 | | | |
| 2 | 1,492 | | | |
| 3 | 1,599 | | | |
| 4 | 1,650 | | | |
| 5 | 2,079 | | | |
| 6 | 903 | | | |
| 7 | 750 | | | |
| 8 | 2,675 | | | |
| 9 | 635 | | | |
| 10 | 1,119 | | | |
| 11 | 1,234 | | | |
| 12 | 1,735 | 1,735 | 1,735 | 1,735 |
| 13 | | 1,305 | 232 | 2,378 |
| 14 | | 1,276 | 203 | 2,349 |
| 15 | | 1,246 | 173 | 2,319 |
| 16 | | 1,217 | 144 | 2,290 |
| 17 | | 1,188 | 115 | 2,261 |
| 18 | | 1,159 | 86 | 2,232 |
| 19 | | 1,130 | 57 | 2,203 |
| 20 | | 1,101 | 27 | 2,174 |
| 21 | | 1,071 | -2 | 2,145 |
| 22 | | 1,042 | -31 | 2,115 |
| 23 | | 1,013 | -60 | 2,086 |
| 24 | | 984 | -90 | 2,057 |

| A خط زمني | B الرحلات السياحية | C التنبؤ (الرحلات السياحية) | D الحد الأدنى للثقة | E الحد الأعلى للثقة |
|---|---|---|---|---|
| 1 | 1,653 | | | |
| 2 | 1,492 | | | |
| 3 | 1,599 | | | |
| 4 | 1,650 | | | |
| 5 | 2,079 | | | |
| 6 | 903 | | | |
| 7 | 750 | | | |
| 8 | 2,675 | | | |
| 9 | 635 | | | |
| 10 | 1,119 | | | |
| 11 | 1,234 | | | |
| 12 | 1,735 | 1,735 | 1,735 | 1,735 |
| 13 | | 1,305 | 232 | 2,378 |
| 14 | | 1,276 | 203 | 2,349 |
| 15 | | 1,246 | 173 | 2,319 |
| 16 | | 1,217 | 144 | 2,290 |
| 17 | | 1,188 | 115 | 2,261 |
| 18 | | 1,159 | 86 | 2,232 |




شكل 4.15: تغيير أسماء الأعمدة



كما تُلاحظ في المخطط الخطي الذي تم إنشاؤه سابقًا، فإن تسمية المحور الأفقي ليست مناسبة، حيث تظهر سلسلة من الأرقام كعنوان للمحور الأفقي بدلًا من أسماء الشهور.

ولتصحيح ذلك، يجب تعديل المُخطط عن طريق تغيير سلسلة البيانات.

## لتعديل سلسلة البيانات:

< في ورقة عمل "2019"، وفي نهاية العمود B، أضف أسماء الأشهر للعام 2023. 1

< في ورقة العمل التي تحتوي على قيم التوقُّع المُنشأة والمُخطط، اضغط داخل المُخطط، 2 ثم اضغط على أيقونة **عوامل تصفية المخطط** (Chart Filters). 3

< اضغط على **تحديد البيانات** (Select data). 4

< ستظهر نافذة **تحديد مصدر البيانات** (Select data source)، ومن **تسميات المحور (الفئة) الأفقي** (Horizontal (Category) Axis Labels)، اضغط على **تحرير** (Edit). 5

< من ورقة عمل "2019"، حدِّد أسماء الشهور. 6

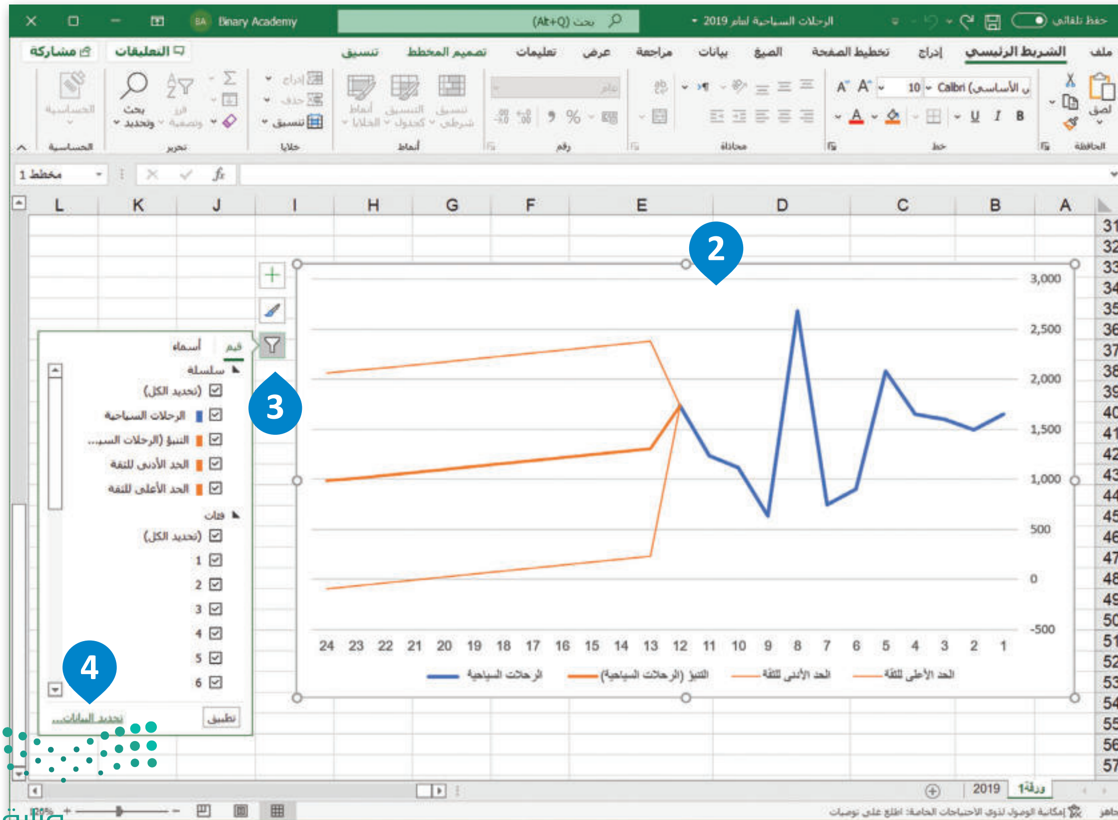
< من نافذة **تسميات المحاور** (Axis Labels)، اضغط على **موافق** (OK). 7

< ثم من نافذة **تحديد مصدر البيانات** (Select Data Source)، اضغط على **موافق** (OK). 8

< ستظهر أسماء الشهور كتسميات على المحور الأفقي. 9

| | D | C | B | A |
|---|---|---|---|---|
| 1 | | الرحلات السياحية | الشهر | |
| 2 | | 1,653 | يناير 2019 | 1 |
| 3 | | 1,492 | فبراير 2019 | 2 |
| 4 | | 1,599 | مارس 2019 | 3 |
| 5 | | 1,650 | أبريل 2019 | 4 |
| 6 | | 2,079 | مايو 2019 | 5 |
| 7 | | 903 | يونيو 2019 | 6 |
| 8 | | 750 | يوليو 2019 | 7 |
| 9 | | 2,675 | أغسطس 2019 | 8 |
| 10 | | 635 | سبتمبر 2019 | 9 |
| 11 | | 1,119 | أكتوبر 2019 | 10 |
| 12 | | 1,234 | نوفمبر 2019 | 11 |
| 13 | | 1,735 | ديسمبر 2019 | 12 |
| 14 | | | يناير 2023 | 13 |
| 15 | | | فبراير 2023 | 14 |
| 16 | | | مارس 2023 | 15 |
| 17 | | | أبريل 2023 | 16 |
| 18 | | | مايو 2023 | 17 |
| 19 | | | يونيو 2023 | 18 |
| 20 | | | يوليو 2023 | 19 |
| 21 | | | أغسطس 2023 | 20 |
| 22 | | | سبتمبر 2023 | 21 |
| 23 | | | أكتوبر 2023 | 22 |
| 24 | | | نوفمبر 2023 | 23 |
| 25 | | | ديسمبر 2023 | 24 |

1



تحديد مصدر البيانات

نطاق بيانات المخطط:

نطاق البيانات شديد التعقيد لدرجة تعوق عرضه. في حالة تحديد نطاق جديد، سيتم استبدال كافة السلاسل الموجودة في لوحة "السلاسل".

تبديل الصف/العمود

إدخالات وسيلة الإيضاح (سلسلة) | تسميات المحور (الفئة) الأفقي

إضافة | تحرير | إزالة | تحرير

الرحلات السياحية | 1

التنبؤ (الرحلات السياحية) | 2

الحد الأدنى للثقة | 3

الحد الأعلى للثقة | 4

5

الخلايا المخفية والفارغة | موافق | إلغاء الأمر

**5**

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 1,653 |
| 3 | 2 | فبراير 2019 | 1,492 |
| 4 | 3 | مارس 2019 | 1,599 |
| 5 | 4 | أبريل 2019 | 1,650 |
| 6 | 5 | مايو 2019 | 2,079 |
| 7 | 6 | يونيو 2019 | 903 |
| 8 | 7 | يوليو 2019 | 750 |
| 9 | 8 | أغسطس 2019 | 2,675 |
| 10 | 9 | سبتمبر 2019 | 635 |
| 11 | 10 | أكتوبر 2019 | 1,119 |
| 12 | 11 | نوفمبر 2019 | 1,234 |
| 13 | 12 | ديسمبر 2019 | 1,735 |
| 14 | 13 | يناير 2023 | |
| 15 | 14 | فبراير 2023 | |
| 16 | 15 | مارس 2023 | |
| 17 | 16 | أبريل 2023 | |
| 18 | 17 | مايو 2023 | |
| 19 | 18 | يونيو 2023 | |
| 20 | 19 | يوليو 2023 | |
| 21 | 20 | أغسطس 2023 | |
| 22 | 21 | سبتمبر 2023 | |
| 23 | 22 | أكتوبر 2023 | |
| 24 | 23 | نوفمبر 2023 | |
| 25 | 24 | ديسمبر 2023 | |

تسميات المحاور

نطاق تسميات المحور:

='2019'!$B$2:$B$25 = يناير 2019، فب...

موافق | إلغاء الأمر

**7** **6**

تحديد مصدر البيانات

نطاق بيانات المخطط:

نطاق البيانات شديد التعقيد لدرجة تعوق عرضه. في حالة تحديد نطاق جديد، سيتم استبدال كافة السلاسل الموجودة في لوحة "السلاسل".

تبديل الصف/العمود

إدخالات وسيلة الإيضاح (سلسلة) | تسميات المحور (الفئة) الأفقي

إضافة | تحرير | إزالة | تحرير

الرحلات السياحية | يناير 2019

التنبؤ (الرحلات السياحية) | فبراير 2019

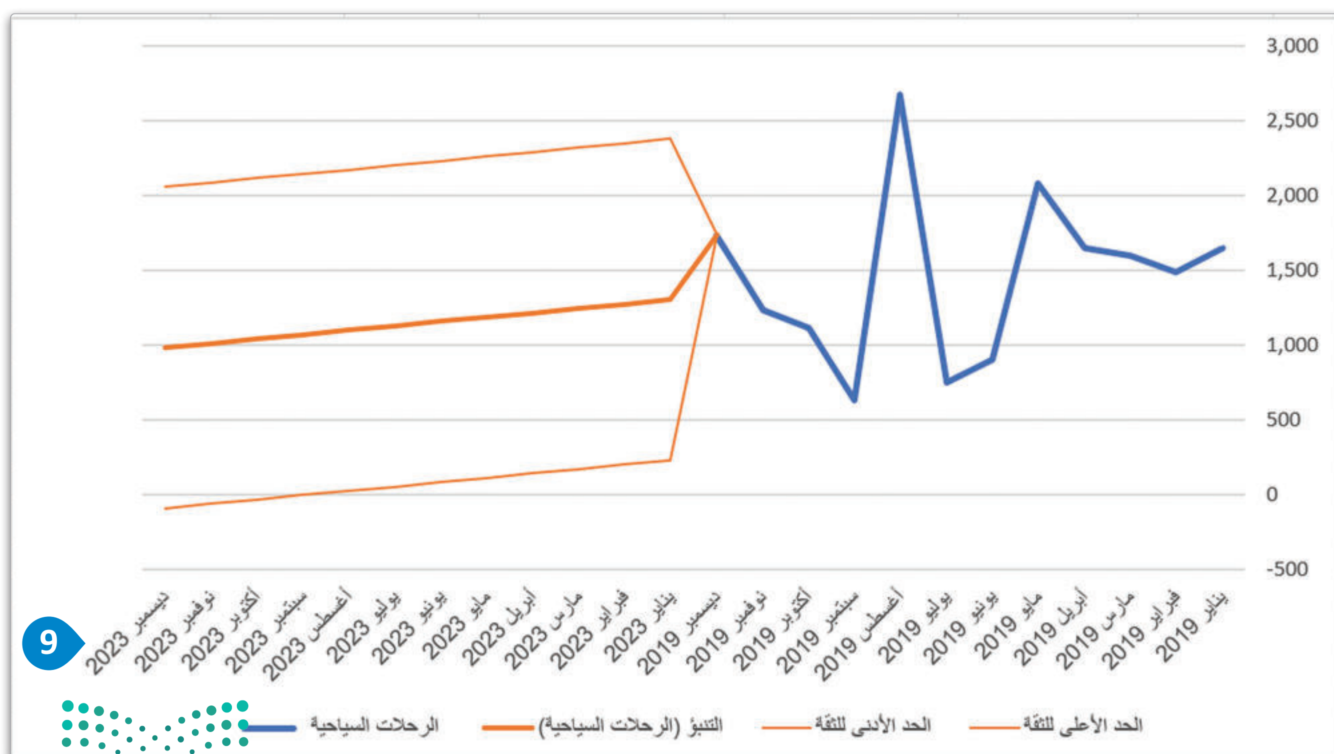
الحد الأدنى للثقة | مارس 2019

الحد الأعلى للثقة | أبريل 2019

مايو 2019

الخلايا المخفية والفارغة | موافق | إلغاء الأمر

**8**




شكل 4.16: تعديل سلسلة البيانات

وللمزيد من التخصيص للمخطط الخطي، يمكنك إضافة عناوين للمحورين x وy.

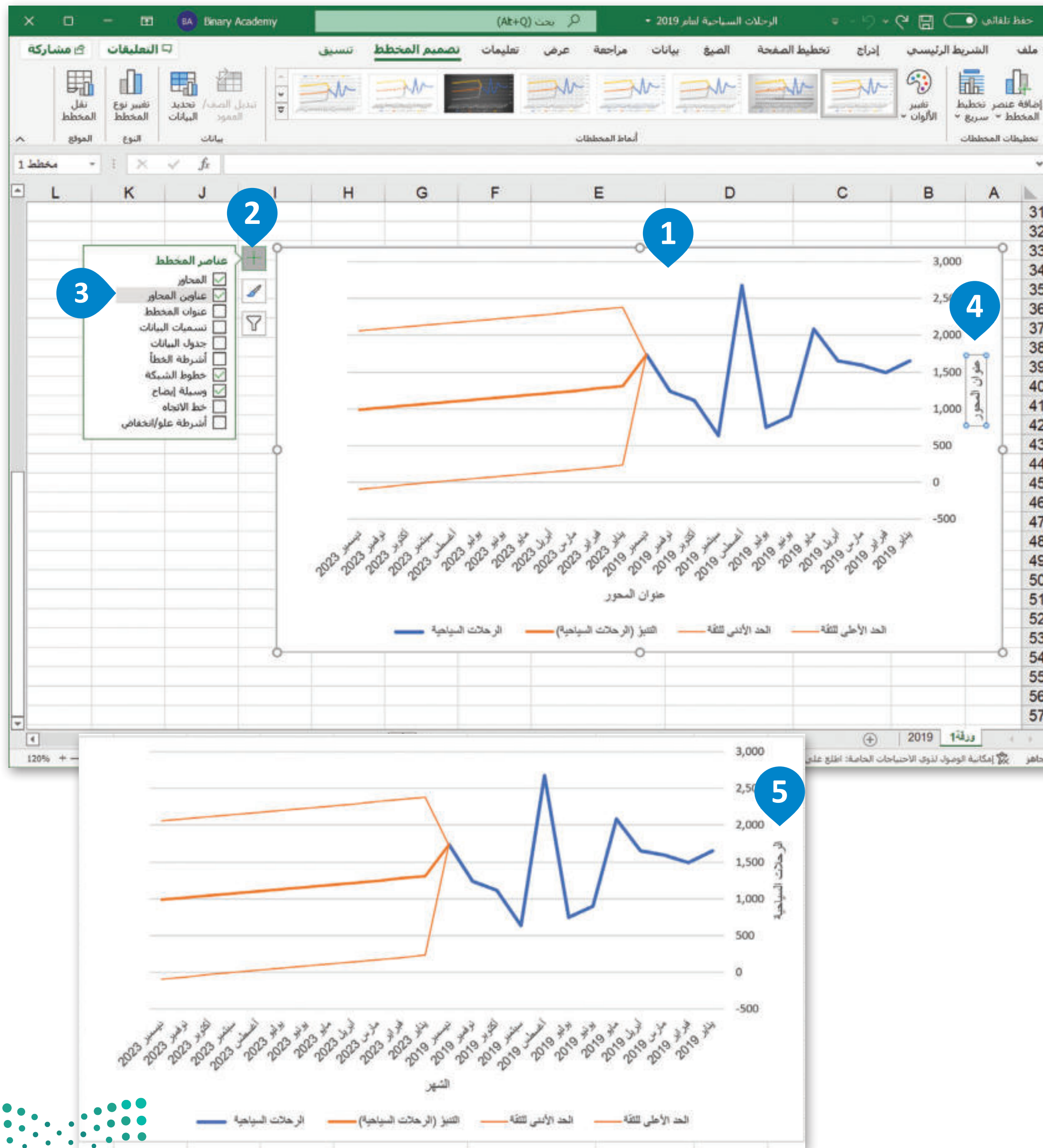
## لتغيير عناوين التسمية:

< حدِّد المخطط الخطي، 1 ثم اضغط على أيقونة +. 2

< حدِّد خيار **عناوين المحاور** (Axis Titles). 3

< في مربعات التسمية التي تظهر، اضغط على كل منها واكتب العنوان الصحيح للمحور. 4

< ستظهر التسميات الصحيحة في المخطط الخطي. 5




شكل 4.17: تغيير عناوين التسمية

## خطوة 6: تحليل البيانات Analyze the data

قد تحتاج في بعض الأحيان إلى الاطِّلاع على مدى الاختلاف بين القيم المُتوقَّعة والقيم الأولية، وذلك من أجل فهم أفضل للظاهرة قيدِ الدراسة وللتوصل إلى الاستنتاجات الصحيحة. بالعودة إلى المثال السابق المتعلق بالسياحة، قد تَودُّ معرفة أيٍّ الشهور في العام 2023 سيشهد ارتفاعًا في عدد الرحلات السياحية الوافدة، وأيها أقل. تُمكِّن هذه المعلومات صنَّاع القرار من وضع الإستراتيجيات التسويقية وتنظيم الحملات الترويجية المناسبة لجذب المزيد من الرحلات السياحية. للحصول على هذه المعلومات، ستطرح قيمة العدد المتوقّع للرحلات السياحية من قيمة العدد السابق للرحلات السياحية، وبالتالي ستحصل على قيمة التباين في البيانات المتوقّعة.

### لإعداد ورقة عمل جديدة:

> في **ورقة 2** (Sheet2) الجديدة، أنشئ عمودًا باسم **الشهر**. 1

> من ورقة عمل "2019"، انسخ القيم الخاصة بالرحلات السياحية وعددها 12، 2 وألصقها في **ورقة 2** (Sheet2)، في عمود **الرحلات السياحية 2019** . 3

> من **ورقة 1** (Sheet 1)، حدِّد القيم المتوقّعة للرحلات السياحية وعددها 12، 4 وألصقها كقيم في **ورقة 2** (Sheet 2)، في عمود **الرحلات السياحية 2023**. 5

**تحليل البيانات (Data Analysis):**

هو الفحص المنهجي للبيانات عن طريق العينات والقياس والتصوير.

| | A | B | C | D |
|---|---|---|---|---|
| 1 | الشهر | الرحلات السياحية 2019 | الرحلات السياحية 2023 | فرق التوقعات |
| 2 | يناير | | | |
| 3 | فبراير | | | |
| 4 | مارس | | | |
| 5 | أبريل | | | |
| 6 | مايو | | | |
| 7 | يونيو | | | |
| 8 | يوليو | | | |
| 9 | أغسطس | | | |
| 10 | سبتمبر | | | |
| 11 | أكتوبر | | | |
| 12 | نوفمبر | | | |
| 13 | ديسمبر | | | |

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 1,653 |
| 3 | 2 | فبراير 2019 | 1,492 |
| 4 | 3 | مارس 2019 | 1,599 |
| 5 | 4 | أبريل 2019 | 1,650 |
| 6 | 5 | مايو 2019 | 2,079 |
| 7 | 6 | يونيو 2019 | 903 |
| 8 | 7 | يوليو 2019 | 750 |
| 9 | 8 | أغسطس 2019 | 2,675 |
| 10 | 9 | سبتمبر 2019 | 635 |
| 11 | 10 | أكتوبر 2019 | 1,119 |
| 12 | 11 | نوفمبر 2019 | 1,234 |
| 13 | 12 | ديسمبر 2019 | 1,735 |
| 14 | 13 | يناير 2023 | |
| 15 | 14 | فبراير 2023 | |
| 16 | 15 | مارس 2023 | |
| 17 | 16 | أبريل 2023 | |
| 18 | 17 | مايو 2023 | |
| 19 | 18 | يونيو 2023 | |
| 20 | 19 | يوليو 2023 | |
| 21 | 20 | أغسطس 2023 | |
| 22 | 21 | سبتمبر 2023 | |

ورقة1 | ورقة2 | 2019

2

| | A | B | C |
|---|---|---|---|
| 1 | الشهر | الرحلات السياحية 2019 | الرحلات السياحية 2023 |
| 2 | يناير | 1,653 | 1,305 |
| 3 | فبراير | 1,492 | 1,276 |
| 4 | مارس | 1,599 | 1,246 |
| 5 | أبريل | 1,650 | 1,217 |
| 6 | مايو | 2,079 | 1,188 |
| 7 | يونيو | 903 | 1,159 |
| 8 | يوليو | 750 | 1,130 |
| 9 | أغسطس | 2,675 | 1,101 |
| 10 | سبتمبر | 635 | 1,071 |
| 11 | أكتوبر | 1,119 | 1,042 |
| 12 | نوفمبر | 1,234 | 1,013 |
| 13 | ديسمبر | 1,735 | 984 |
| 14 | | | |

ورقة1 | ورقة2 | 2019

Select predicted d.

3

5

| خط زمني | الرحلات السياحية | التنبؤ (الرحلات السياحية) | الحد الأدنى للثقة | الحد الأعلى للثقة |
|---|---|---|---|---|
| 1 | 1,653 | | | |
| 2 | 1,492 | | | |
| 3 | 1,599 | | | |
| 4 | 1,650 | | | |
| 5 | 2,079 | | | |
| 6 | 903 | | | |
| 7 | 750 | | | |
| 8 | 2,675 | | | |
| 9 | 635 | | | |
| 10 | 1,119 | | | |
| 11 | 1,234 | | | |
| 12 | 1,735 | 1,735 | 1,735 | 1,735 |
| 13 | | 1,305 | 232 | 2,378 |
| 14 | | 1,276 | 203 | 2,349 |
| 15 | | 1,246 | 173 | 2,319 |
| 16 | | 1,217 | 144 | 2,290 |
| 17 | | 1,188 | 115 | 2,261 |
| 18 | | 1,159 | 86 | 2,232 |
| 19 | | 1,130 | 57 | 2,203 |
| 20 | | 1,101 | [illegible] | 2,174 |
| 21 | | 1,071 | -2 | 2,145 |
| 22 | | 1,042 | -31 | 2,115 |
| 23 | | 1,013 | -60 | 2,086 |
| 24 | | 984 | -90 | 2,057 |

ورقة1 | ورقة2 | 2019

4

شكل 4.18: إعداد ورقة عمل جديدة



وزارة التعليم
Ministry of Education
2022 - 1444

## لحساب فرق التوقُّعات:

< في ورقة 2 (Sheet 2)، أنشئ عمودًا جديدًا باسم **فرق التوقعات** . 1
< في الخلية D2، اكتب الصيغة " C2-B2=". 2
< انسخ الصيغة من الخلية D2 إلى الخلية D13 لحساب باقي القيم. 3

| | A | B | C | D |
|---|---|---|---|---|
| 1 | الشهر | الرحلات السياحية 2019 | الرحلات السياحية 2023 | فرق التوقعات |
| 2 | يناير | 1,653 | 1,305 | =C2−B2 |
| 3 | فبراير | 1,492 | 1,276 | |
| 4 | مارس | 1,599 | 1,246 | |
| 5 | أبريل | 1,650 | 1,217 | |
| 6 | مايو | 2,079 | 1,188 | |
| 7 | يونيو | 903 | 1,159 | |
| 8 | يوليو | 750 | 1,130 | |
| 9 | أغسطس | 2,675 | 1,101 | |
| 10 | سبتمبر | 635 | 1,071 | |
| 11 | أكتوبر | 1,119 | 1,042 | |
| 12 | نوفمبر | 1,234 | 1,013 | |
| 13 | ديسمبر | 1,735 | 984 | |

| | A | B | C | D |
|---|---|---|---|---|
| 1 | الشهر | الرحلات السياحية 2019 | الرحلات السياحية 2023 | فرق التوقعات |
| 2 | يناير | 1,653 | 1,305 | −349 |
| 3 | فبراير | 1,492 | 1,276 | −216 |
| 4 | مارس | 1,599 | 1,246 | −353 |
| 5 | أبريل | 1,650 | 1,217 | −433 |
| 6 | مايو | 2,079 | 1,188 | −891 |
| 7 | يونيو | 903 | 1,159 | 256 |
| 8 | يوليو | 750 | 1,130 | 380 |
| 9 | أغسطس | 2,675 | 1,101 | −1,574 |
| 10 | سبتمبر | 635 | 1,071 | 436 |
| 11 | أكتوبر | 1,119 | 1,042 | −77 |
| 12 | نوفمبر | 1,234 | 1,013 | −221 |
| 13 | ديسمبر | 1,735 | 984 | −751 |
| 14 | | | | |

شكل 4.19: حساب فرق التوقُّعات



## إنشاء مخطط عمودي متفاوت المسافات Creating Clustered Column Chart

بالإضافة إلى مخطط التوقّع الذي قمت بإنشائه سابقاً، ستقوم بإنشاء مخططين آخرين تستخدمهما في التحليل.
بشكل أكثر تحديدًا، ستُنشئ التالي:

< مخطط بياني لتمثيل المقارنة بين بيانات الرحلات السياحية السابقة وبيانات الرحلات السياحية المتوقّعة.

< مخطط بياني يوضح الفرق المتوقَّع بين بيانات الرحلات السياحية السابقة وبيانات الرحلات السياحية المتوقّعة.

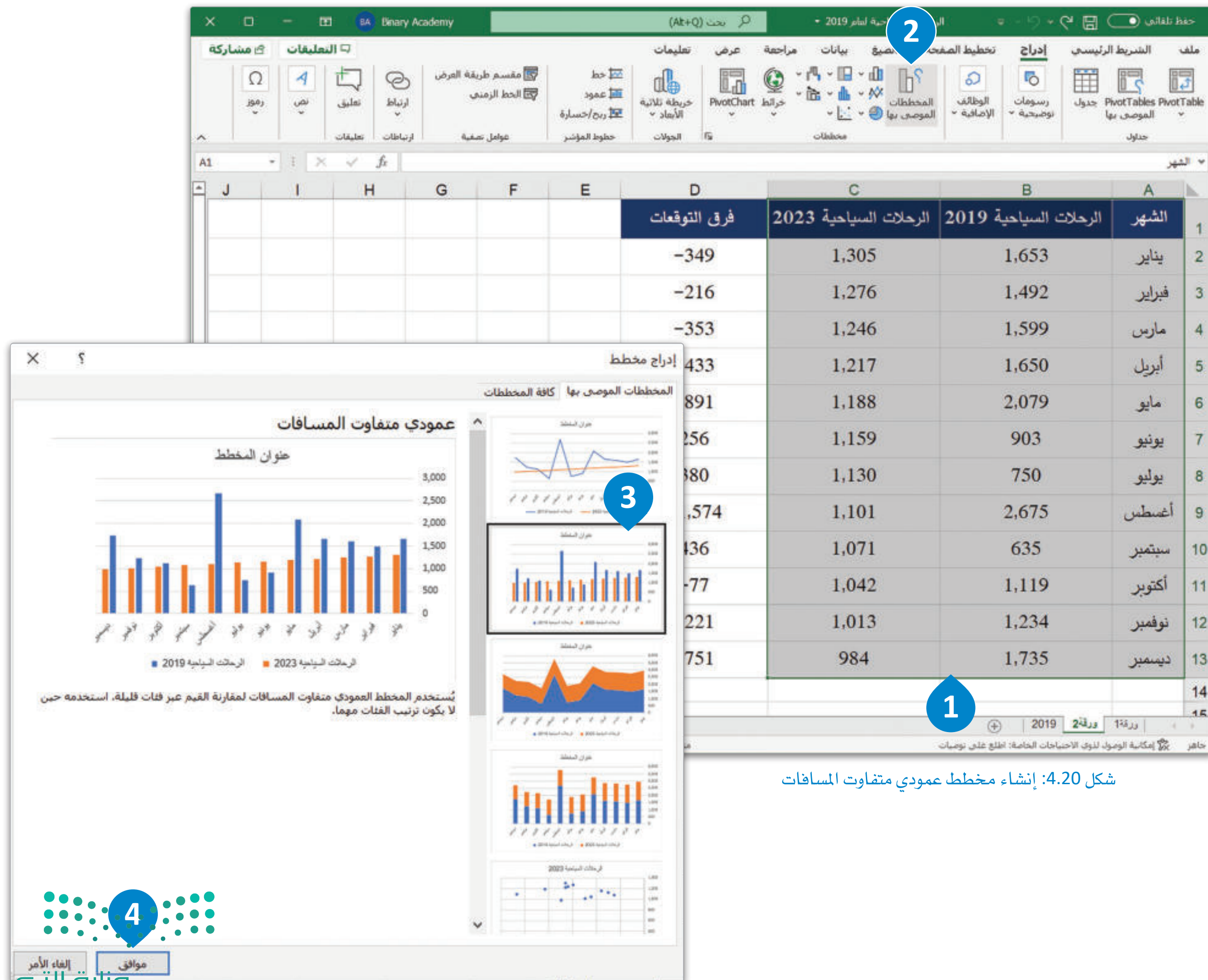
**لإنشاء مخطط عمودي متفاوت المسافات:**

< حدد الأعمدة A و B و C. 1

< من علامة تبويب إدراج (Insert)، اضغط على **المخططات الموصى بها** (Recommended Charts). 2

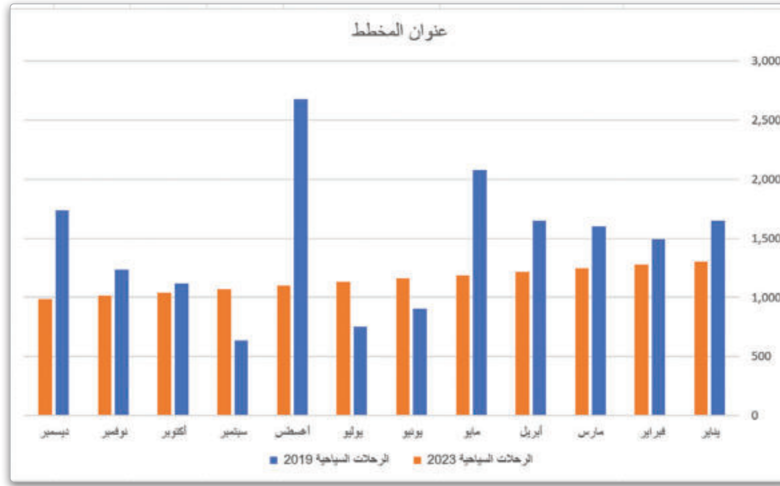
< اختر مخطط **عمودي متفاوت المسافات** (Clustered Column chart). 3

< اضغط على **موافق** (OK). 4

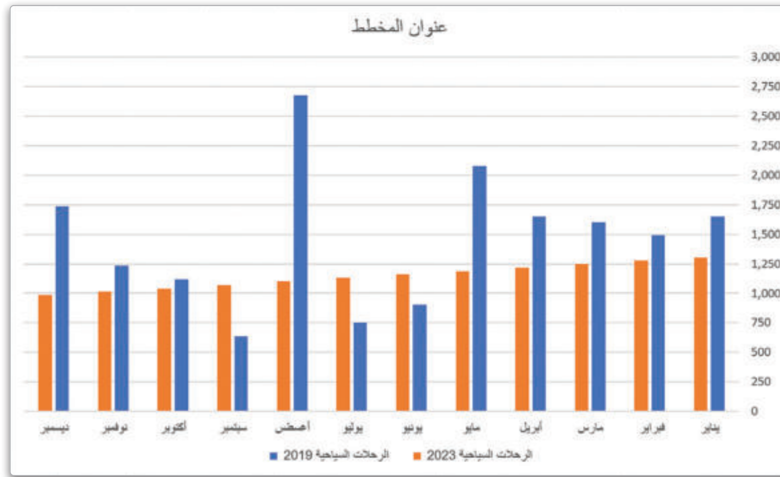



شكل 4.20: إنشاء مخطط عمودي متفاوت المسافات

يمثّل هذا المخطط البياني المقارنة بين بيانات الرحلات السياحية السابقة، وبيانات الرحلات السياحية المتوقّعة. ويمكنك تغيير حدود أو وحدات المحور العمودي لعرض تمثيلٍ أوضح للبيانات.



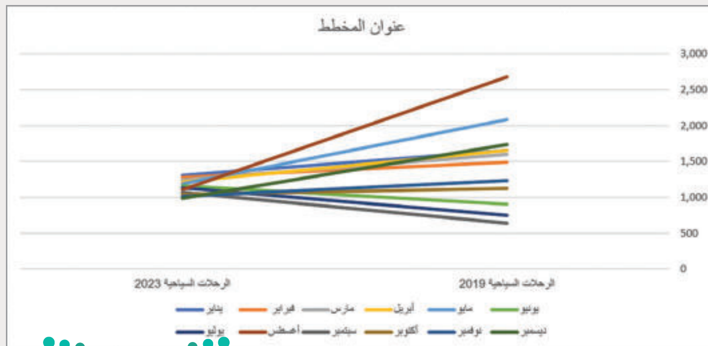

عند المقارنة مع المخطط البياني الأول، يحتوي هذا المخطط على تَدرُّج أصغر للوحدات بقيمة 250 (بدلًا من 500 كما في المخطط البياني الأولي).




شكل 4.21: مخطط عمودي متفاوت المسافات

عندما يتعلق الأمر بتمثيل المعلومات في إكسل، فمن المهم جدًا اختيار النوع المناسب والصحيح للمخططات، وذلك ليتمكن الجمهور من قراءتها وفهمها بسهولة.

ولهذا الغرض، عليك اختيار المخططات التي يُوصي بها إكسل، حيث تكون في أغلب الحالات مناسبة لنوع البيانات المتوفر، فاختيارك مخططًا غير مناسب لتمثيل البيانات سيؤدي إلى الحصول على تمثيل مرئي لا يمكن فهمه أو تفسير محتواه. يُظهر الشكل 4.22 مثالًا على نتيجة الاختيار الخاطئ لنوع المخطط البياني.

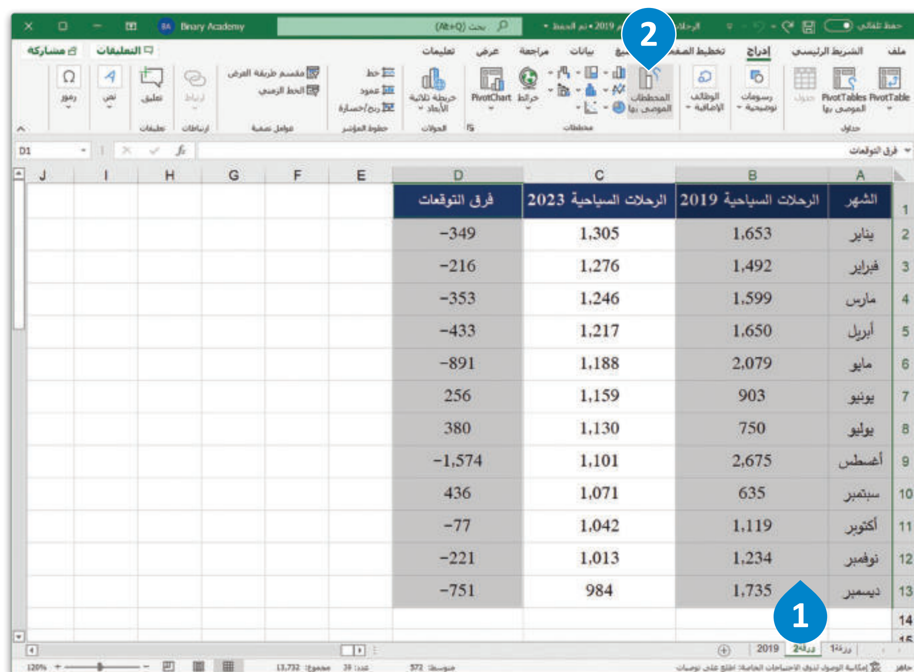



شكل 4.22: مخطط خطي غير مفهوم

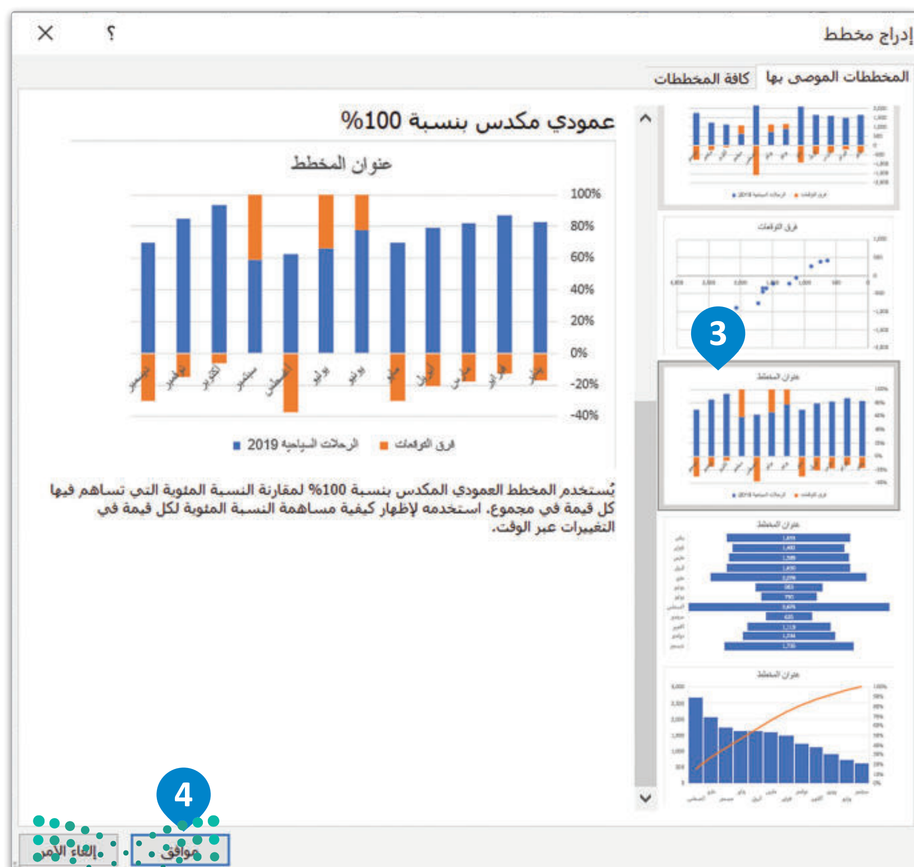
## المخطط العمودي المكدس Stacked Column Chart

**لإنشاء مخطط عمودي مكدس:**

- حَدِّد الأعمدة A وB وD. 1
- من علاقة تبويب إدراج (Insert)، اضغط على المخططات الموصى بها (Recommended Charts). 2
- اختر مخطط عمودي مكدس بنسبة 100 % (Stacked Column). 3
- اضغط على موافق (OK). 4



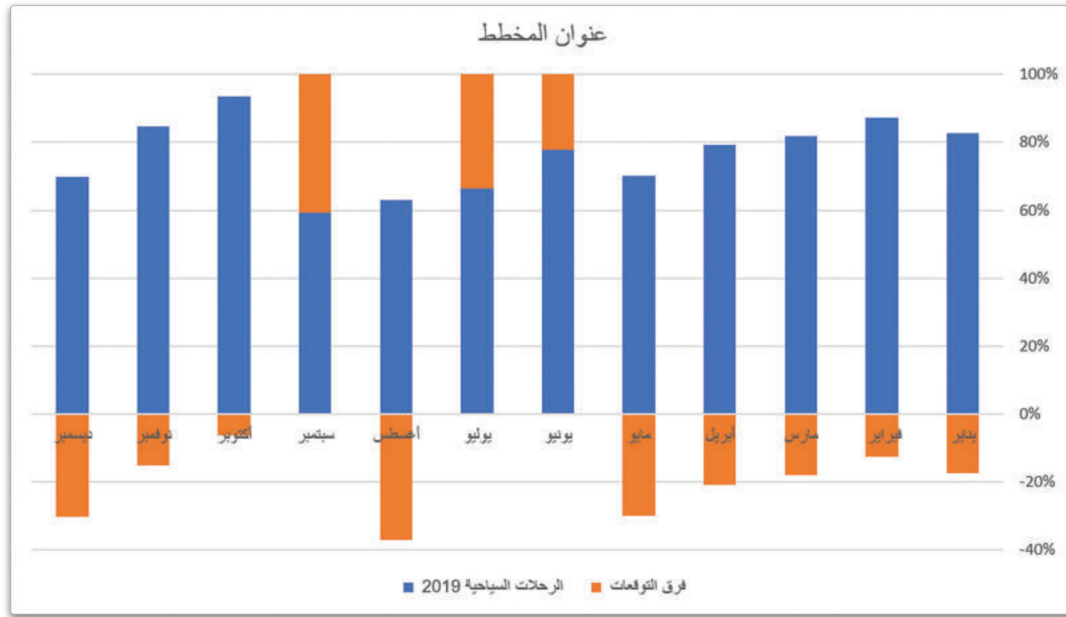
| الشهر | الرحلات السياحية 2019 | الرحلات السياحية 2023 | فرق التوقعات |
|---|---|---|---|
| يناير | 1,653 | 1,305 | -349 |
| فبراير | 1,492 | 1,276 | -216 |
| مارس | 1,599 | 1,246 | -353 |
| أبريل | 1,650 | 1,217 | -433 |
| مايو | 2,079 | 1,188 | -891 |
| يونيو | 903 | 1,159 | 256 |
| يوليو | 750 | 1,130 | 380 |
| أغسطس | 2,675 | 1,101 | -1,574 |
| سبتمبر | 635 | 1,071 | 436 |
| أكتوبر | 1,119 | 1,042 | -77 |
| نوفمبر | 1,234 | 1,013 | -221 |
| ديسمبر | 1,735 | 984 | -751 |




شكل 4.23: إنشاء مخطط عمودي مكدس

يُستخدم المخطط العمودي المكدس لمقارنة قيمة معينة مع مجموع قيم تتضمن تلك القيمة، ويمكنك استخدامه لإظهار كيفية مساهمة قيمة معينة في التغييرات عبر الوقت.

يُمثل المخطط البياني في الشكل 4.24 الفرق المتوقّع بين عدد الرحلات السياحية السابقة وعدد الرحلات السياحية المتوقّعة، ويمكنك تغيير حدود أو وحدات المحور الرأسي لتمثيل المعلومات بشكل أكثر وضوحًا.

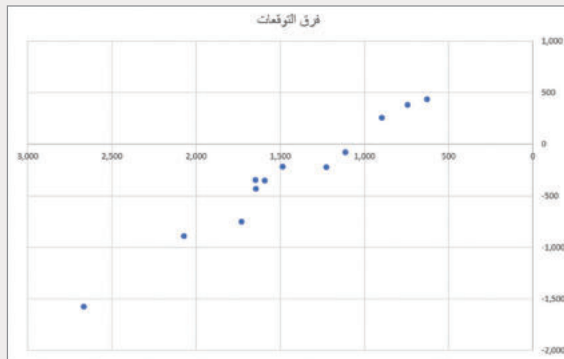


شكل 4.24: مخطط عمودي مكدّس

قد لا تكون المخططات المُوصى بها في إكسل مناسبةً لتمثيل المعلومات في حالات معينة. فعلى سبيل المثال، يُوصي إكسل في المثال السابق باستخدام **مخطط قُمعي** (Funnel Chart) كما في الشكل 4.25، أو **مخطط مبعثر** (Scatter Chart) كما في الشكل 4.26، وإذا قمت باتِّباع التوصية واستخدام تلك المخططات، فستحصل على مخططات غير مناسبة، ويصعب قراءتها أو فهم معانيها.



شكل 4.25: مخطط قُمعي



شكل 4.26: مخطط مبعثر

يجب الأخذ بالاعتبار أن بعض أنواع المخططات قد لا تتناسب مع حاجات أو قدرات الجمهور، فبعض تلك الأنواع – وكما هو الحال في بعض البرامج والأدوات – هو مناسبٌ للعلماء والباحثين وخبراء التسويق. ومن الشائع أيضًا أن المسؤولين التنفيذيين ومدراء الشركات يحتاجون أنواعاً محددة من المخططات.



# تمرينات

1

| حدد الجملة الصحيحة والجملة الخاطئة فيما يلي: | صحيحة | خاطئة |
|---|---|---|
| 1. يُعَدُّ التوقُّع الطريقة الوحيدة للتنبؤ بالبيانات المستقبلية بناءً على البيانات السابقة. | ⚪ | ⚪ |
| 2. إن تحديد الخطوات بوضوح قبل تنفيذ إجراء التوقُّع هو أهم متطلبات التوقُّع الجيد. | ⚪ | ⚪ |
| 3. ليس من الضروري اتباع الخطوات الستّ للتوقُّع بالترتيب. | ⚪ | ⚪ |
| 4. يمنحك نطاق الثقة معلومات حول نسبة الشك وعدم التيقن بحدوث التنبؤ. | ⚪ | ⚪ |
| 5. التنبؤ بنطاق ثقة يساوي 95% هو أكثر دقة من تنبؤ بنطاق ثقة يساوي 75%. | ⚪ | ⚪ |
| 6. يشير التنبؤ والتوقُّع إلى نفس الإجراء. | ⚪ | ⚪ |
| 7. يُفَضّل استخدام المخطط البياني الخطي على المخطط العمودي في تمثيل التوقُّع. | ⚪ | ⚪ |
| 8. يُفَضَّل اختيار المخططات التي يوصي بها إكسل لمناسبتها لنوع البيانات المستخدم. | ⚪ | ⚪ |
| 9. يعتمد اختيار المخطط الصحيح لتمثيل البيانات على المعلومات التي سيتم تمثيلها بالمخطط. | ⚪ | ⚪ |
| 10. تحدد حدود الثقة الدنيا والعليا النطاق المقبول لقيم البيانات. | ⚪ | ⚪ |



**2** قارن بين نوعي المخططين اللذين يقدمهما إجراء التوقُّع في إكسل، ولماذا يُعدُّ المخطط الخطي هو الأنسب حسب رأيك؟

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

**3** قم بزيارة صفحة بيانات السياحة التابعة لوزارة السياحة السعودية في البوابة الوطنية للبيانات المفتوحة (https://data.gov.sa) ثم قم بتحميل بيانات الرحلات السياحية للعامين 2017 و 2018. ثم طبِّق إجراء التوقُّع للعام 2023 ووضِّح النتائج. اذكر المخططات البيانية التي ستختارها لعرض النتائج؟ ولماذا؟

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________



**4** حلّل نتائج التوقُّع لبيانات الرحلات السياحية التي حصلت عليها من التمرين الثالث للعامين 2017 و2018، ثم قم بإنشاء أوراق العمل المطلوبة في إكسل وقارن بين النتائج. ما هو الجزء المهم في هذه المخططات التوضيحية عند إجراء المقارنة بينها؟ وضِّح إجابتك.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**5** مستعينًا بأوراق عمل إكسل التي أنشأتها في التمرين الرابع، قم بتمثيل البيانات باستخدام الأنواع المختلفة للمخططات في إكسل. ما هي أنسب أنواع المخططات لتمثيل هذه المعلومات؟ وضِّح إجابتك.

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________

______________________________________________






الدرس الثالث

# التحسين (Optimization)

## التعامل مع مشاكل التحسين
## Dealing with Optimization Problems

يواجه الإنسان المشاكل المتعلقة بالتحسين في الكثير من جوانب الحياة ومجالاتها كالأعمال التجارية، والاقتصاد، والرياضيات، والهندسة، والعلوم. تَكمُن المشكلة الأساسية في محاولة إيجاد الطريقة المُثلى أو الأكثر فاعلية لاستخدام الموارد المحدودة لتحقيق هدف معين والذي قد يكون زيادة الربح، أو تقليل التكلفة، أو تقليل الوقت اللازم لتنفيذ مشروع ما، أو اختصار الوقت أو المسافة للسفرِ إلى وجهة معينة.

يُعدُّ استقطاب المزيد من الرحلات السياحية إلى بلد ما أحد أهداف التحسين، كما يمكن النظر إلى بعض الأمثلة الأُخرى حول التحسين، كتحديد الميزانية المناسبة لحملة تسويقية، أو تصميم الجدول المثالي لعمل الموظفين، أو خفض تكلفة التوصيل.

**التحسين (Optimization):**
هو عملية اختيار العنصر الأفضل من مجموعة من البدائل وذلك تحت قيود معينة.

## ما هي أداة إكسل سولفر؟ What is Excel Solver?

تُعتبر أداة **إكسل سولفر** (Excel Solver) أداة برمجية تُستخدم لمحاكاة وتحسين نماذج الهندسة والأعمال المختلفة. تنتمي هذه الأداة إلى مجموعة خاصة من أدوات إكسل يُشار إليها باسم **أدوات تحليل ماذا-إذا** (What-If Analysis tools) وتستخدم للوصول إلى أفضل الحلول لنموذجٍ يحتوي على مُدخلات متعددة.

تُعدُّ الحاجة إلى تحديد قيمة لخلية واحدة (تُسمى الخلية الهدف) عن طريق تغيير قيم خلايا أخرى معينة (تسمى خلايا المُتغير) وذلك باستخدام القيود أو بدونها، من أكثر الاستخدامات شيوعًا لأداة إكسل سولفر، حيث تعتبر هذه الأداة مثالية لحل مشاكل البرمجة الخطية (المعروفة أيضًا باسم مشاكل التحسين الخطي)، ولذلك يُطلق عليها أحيانًا اسم **البرمجة الخطية لسولفر** (Linear Programming Solver).

تُعد أداة إكسل سولفر مثالية للتعامل مع مشاكل التحسين، حيث تُمكِّنك من استخدام جداول البيانات لإدراج متغيرات القرار وقيود النموذج، ثم تنفيذ دالة الهدف التي تصفها. وفي حين يُعدُّ استخدام **النماذج الشبكية** (Graphical Models) مثالياً لحل النموذج الذي يحتوي على متغيرين للقرار، تتضمن الكثير من المشاكل في الحياة الواقعية متغيرات عديدة، ولذلك هناك حاجة إلى استخدام تقنيات وحسابات مُعقدة للوصول إلى الحل الأمثل لمثل تلك المشاكل.

تُساهم جداول البيانات وأداة سولفر في حل مشاكل التحسين بطريقة مُبسطة وملائمة للمستخدمين بِصرف النظر عن قدراتهم الحسابية.

**من الضروري فهم إمكانية إجراء تغييرات على مُعاملات أداة سولفر أو على الإجراء الذي يتم تنفيذه في أي وقت، وبذلك بناءً على النتائج التي تقدمها الأداة، يمكن إعادة تقييم المشكلة وتحديد ما إذا كانت هناك حاجة إلى تنفيذ عملية هذه الأداة مع تطبيق قيود معينة. لا تُعدُّ النتائج في سولفر مجرد أرقام فقط، بل هي قيم ذات معنى محدد تخص الموقف قيد الدراسة، ولذلك يجب على عالم البيانات أو محلل الأعمال إجراء تقييمٍ نقدي لهذه النتائج واتخاذ إجراءات إضافية إذا كانت النتائج غير مرضية.**



## صياغة المشكلة Formulating the Problem

قبل تشغيل أداة إكسل سولفر، يجب صياغة المشكلة (النموذج) في ورقة عمل، حيث يعرّف هذا النموذج المشكلة التي تريد إيجاد حل لها.

لقد جمعت في الدرس السابق بيانات الرحلات السياحية الشهرية للعام 2019 من صفحة بيانات السياحة التابعة لوزارة السياحة السعودية في البوابة الوطنية للبيانات المفتوحة (https://data.gov.sa)، وفي هذا الدرس ستستخدم نفس البيانات من أجل الوصول إلى نتائج محددة تساعد في تنظيم حملات للترويج للسياحة في المملكة العربية السعودية للعام 2023، وذلك باستخدام أداة إكسل سولفر. لتحقيق ذلك، يتعين عليك أولًا صياغة المشكلة ثم تحديد نوع المعلومات التي تريد الحصول عليها من أداة إكسل سولفر.

**تعريف المشكلة**

لو افترضت أنك وكيل سفر وأردت تنظيم حملة ترويجية للسياحة للعام 2023 في المملكة العربية السعودية، فمن المُهم في هذه الحالة معرفة عدد الرحلات السياحية التي تمت خلال العام 2019، وبناءً على ذلك ستستطيع تحديد الطرق والوسائل التي يمكن الاستعانة بها لجذب المزيد من السياح في العام 2023.

**جمع البيانات**

سيتم الحصول على بيانات الرحلات السياحية الشهرية المطلوبة لعام 2019 من صفحة بيانات السياحة التابعة لوزارة السياحة السعودية في البوابة الوطنية للبيانات المفتوحة (https://data.gov.sa).

**تحديد الأهداف**

انطلاقاً من حقيقة أنه في العام 2019 وصل عدد الرحلات السياحية للمملكة العربية السعودية إلى 17,526,000، يمكن تحديد الهدف بزيادة هذا العدد إلى 25,000,000 رحلة سياحية في العام 2023.

**تحديد المتغيرات المتأثرة**

في البداية ستقوم بتقييم بيانات الرحلات السياحية الشهرية لعام 2019 لتحديد نقاط الضعف أو الإشكاليات الكامنة في البيانات المتعلقة بكل شهر من شهور هذا العام، وبناءً على هذا التقييم الأولي، ستُحدد المتغير الذي يجب تغييره من أجل تحقيق الهدف، ثم بناءً على النتائج ستُقرر ما إذا كان يجب تنفيذ أداة سولفر بقيود أم لا.

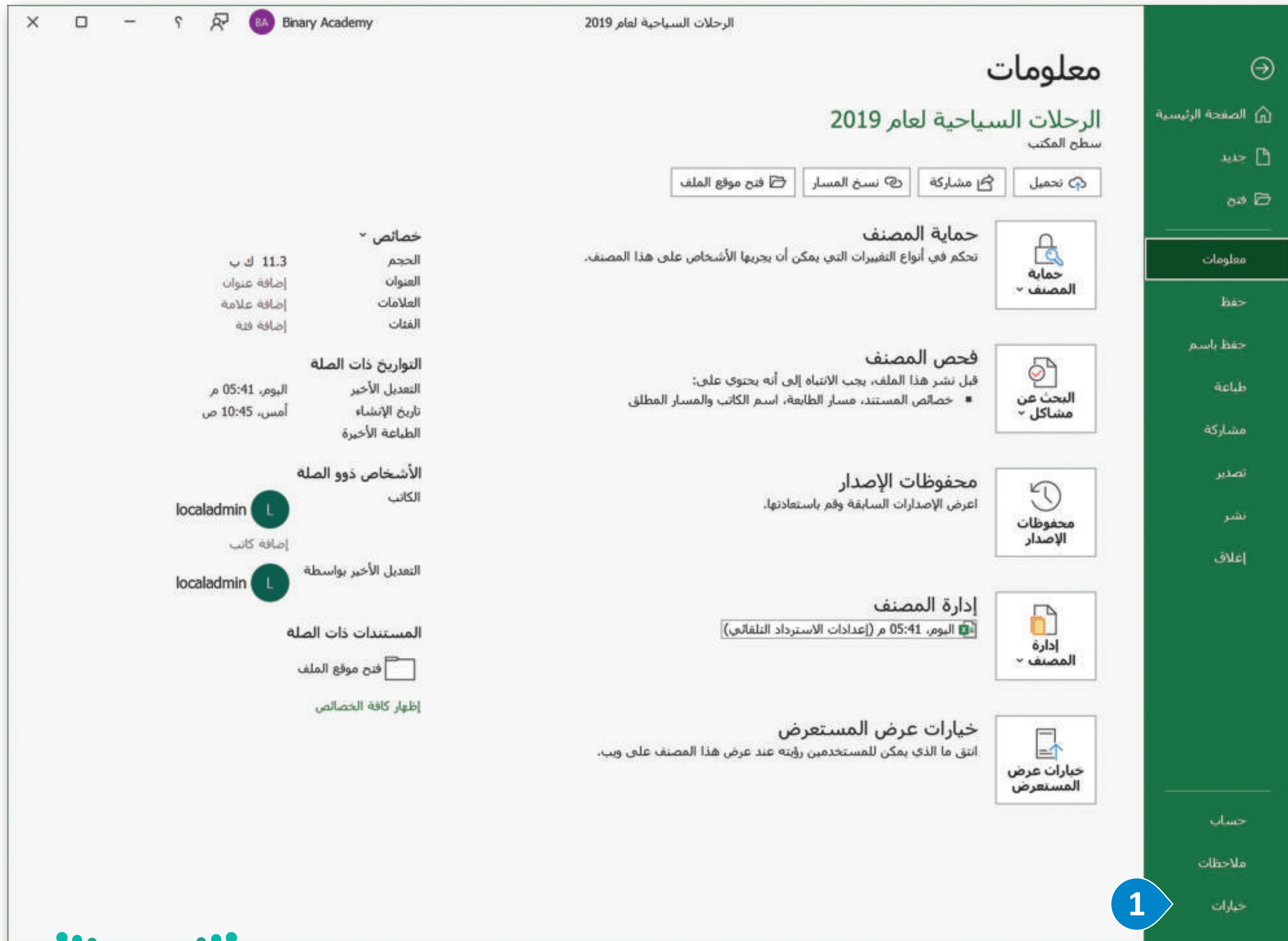


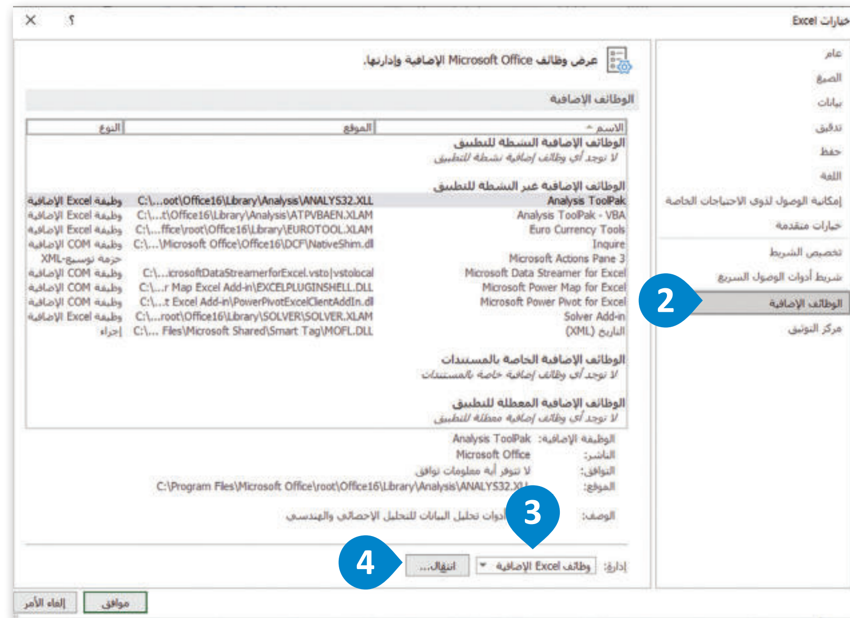
## أداة إكسل سولفر الإضافية The Excel Solver Add-In

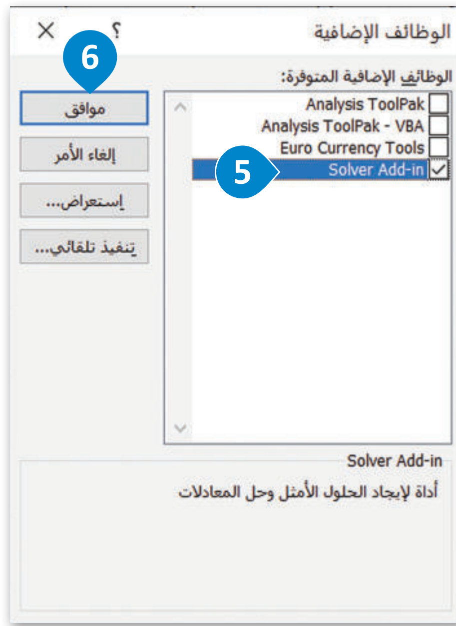
الآن وبعد صياغة المشكلة بشكل جيد، يمكنك فتح برنامج إكسل وتشغيل أداة إكسل سولفر. يجب عليك أولًا تنشيط هذه الأداة من نافذة **خيارات إكسل** (Excel Options)، لأنه لا يتم تنشيطها تلقائيًا عند تثبيت **مايكروسوفت أوفيس** (Microsoft Office).

**لتنشيط أداة إكسل سولفر الإضافية (Excel Solver Add-In):**

- من علامة تبويب **ملف** (File)، اضغط على **خيارات** (Options). 1
- من نافذة **خيارات** Excel (Excel Options)، اضغط على **الوظائف الإضافية** (Add-Ins). 2
- من مربع **إدارة** (Manage)، اختر **وظائف** Excel **الإضافية** (Excel Add-Ins)، 3 ثم اضغط **انتقال** (GO). 4
- من نافذة **الوظائف الإضافية** (Add-Ins)، حدِّد Solver Add-In (الأداة الإضافية سولفر). 5
- اضغط على **موافق** (OK). 6
- ستظهر أيقونة أداة Solver (سولفر). 7






| | الشهر | الرحلات السياحية |
|---|---|---|
| 1 | يناير 2019 | 1,653 |
| 2 | فبراير 2019 | 1,492 |
| 3 | مارس 2019 | 1,599 |
| 4 | أبريل 2019 | 1,650 |
| 5 | مايو 2019 | 2,079 |
| 6 | يونيو 2019 | 903 |
| 7 | يوليو 2019 | 750 |
| 8 | أغسطس 2019 | 2,675 |
| 9 | سبتمبر 2019 | 635 |
| 10 | أكتوبر 2019 | 1,119 |
| 11 | نوفمبر 2019 | 1,234 |
| 12 | ديسمبر 2019 | 1,735 |

عدد الرحلات السياحية الوافدة للمملكة العربية السعودية (مقدرًا بالآلاف).

7 Solver

شكل 4.27: تنشيط الأداة الإضافية إكسل سولفر



وزارة التعليم
Ministry of Education
2022 - 1444

## استخدام أداة سولفر Using Solver

عليك قبل كل شيء حساب العدد الإجمالي للرحلات السياحية لعام 2019 (مقدرًا بالآلاف)، وللقيام بذلك ستستخدم دالة **الجمع** (SUM) في إكسل، حيث ستقوم بتحديد وإيجاد مجموع الخلايا التي تحتوي على أعداد الرحلات السياحية الشهرية.

**لحساب عدد الرحلات الإجمالي:**

- افتح ملف إكسل "الرحلات السياحية لعام 2019". 1
- في الخلية B14 اكتب "الإجمالي". 2
- في الخلية C14 اكتب "=SUM(C2:C13)" 3 واضغط على **Enter ↵**.
- سيظهر المجموع الإجمالي في الخلية. 4

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 1,653 |
| 3 | 2 | فبراير 2019 | 1,492 |
| 4 | 3 | مارس 2019 | 1,599 |
| 5 | 4 | أبريل 2019 | 1,650 |
| 6 | 5 | مايو 2019 | 2,079 |
| 7 | 6 | يونيو 2019 | 903 |
| 8 | 7 | يوليو 2019 | 750 |
| 9 | 8 | أغسطس 2019 | 2,675 |
| 10 | 9 | سبتمبر 2019 | 635 |
| 11 | 10 | أكتوبر 2019 | 1,119 |
| 12 | 11 | نوفمبر 2019 | 1,234 |
| 13 | 12 | ديسمبر 2019 | 1,735 |
| 14 | | الإجمالي | =SUM(C2:C13) |

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 1,653 |
| 3 | 2 | فبراير 2019 | 1,492 |
| 4 | 3 | مارس 2019 | 1,599 |
| 5 | 4 | أبريل 2019 | 1,650 |
| 6 | 5 | مايو 2019 | 2,079 |
| 7 | 6 | يونيو 2019 | 903 |
| 8 | 7 | يوليو 2019 | 750 |
| 9 | 8 | أغسطس 2019 | 2,675 |
| 10 | 9 | سبتمبر 2019 | 635 |
| 11 | 10 | أكتوبر 2019 | 1,119 |
| 12 | 11 | نوفمبر 2019 | 1,234 |
| 13 | 12 | ديسمبر 2019 | 1,735 |
| 14 | | الإجمالي | 17,526 |

شكل 4.28: حساب العدد الإجمالي للرحلات السياحية لعام 2019

عليك بعدَ ذلك فتح ملف إكسل الذي يحتوي على بيانات الرحلات السياحية لعام 2019 في ورقة عمل بإسم "2019"، ثم ستُنشئ ورقة عمل جديدة باسم "Solver" بنفس البيانات الموجودة في ورقة عمل "2019".

من المهم القيام بالخطوة المذكورة نظرًا لأن استخدام أداة سولفر يؤدي إلى تغيير القيم الموجودة في البيانات بشكل دائم وبدون خيار للتراجع، لذلك عليك الاحتفاظ بورقة عمل البيانات الأصلية (الورقة "2019" في هذا المثال)، لاستخدامها من جديد في حالة كانت نتائج أداة سولفر غير مُرضية.

إن أسهل طريقة لنسخ القيم الموجودة في ورقة العمل "2019" هي بإنشاء نسخة من الورقة نفسها.

## لنسخ ورقة عمل في إكسل:

< فتح ملف إكسل "الرحلات السياحية لعام 2019". 1

< اضغط بزر الفأرة الأيمن على ورقة عمل "2019"، 2 وحدِّد **نقل أو نسخ** (Move or Copy). 3

< في نافذة النقل أو النسخ التي ستظهر، حدِّد "2019" 4 ثم حدد خيار **إنشاء نسخة** (Create a copy). 5

< اضغط على **موافق** (OK). 6

< تم إنشاء الورقة الجديدة. 7

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 1,653 |
| 3 | 2 | فبراير 2019 | 1,492 |
| 4 | 3 | مارس 2019 | 1,599 |
| 5 | 4 | أبريل 2019 | 1,650 |
| 6 | 5 | مايو 2019 | 2,079 |
| 7 | 6 | يونيو 2019 | 903 |
| 8 | 7 | يوليو 2019 | 750 |
| 9 | 8 | أغسطس 2019 | 2,675 |
| 10 | 9 | سبتمبر 2019 | 635 |
| 11 | 10 | أكتوبر 2019 | 1,119 |
| 12 | 11 | نوفمبر 2019 | 1,234 |
| 13 | 12 | ديسمبر 2019 | 1,735 |
| 14 | | الإجمالي | 17,526 |

نقل أو نسخ

نقل الأوراق المحددة

إلى المصنف:

الرحلات السياحية لعام 2019.xlsx

قبل الورقة:

2019

(تحريك إلى النهاية)

إنشاء نسخة

موافق | إلغاء الأمر

شكل 4.29: نسخ ورقة عمل في إكسل

## البحث عن الخلايا ذات الإشكاليات الكامنة Find the problematic cell values

كما ذُكر سابقًا، سيتم تقييم بيانات الرحلات السياحية الشهرية لعام 2019 من أجل إيجاد القيم ذات الإشكاليات (إن وُجدت)، ثم تحديد المتغيرات التي تحتاج إلى تعديل لتحقيق الهدف المنشود. يتم تعريف هذه القيم في هذا المثال بالقيم التي تحتوي على العدد الأقل للرحلات السياحية. إذا ألقيت نظرة فاحصة على ورقة عمل إكسل "2019"، ستُلاحظ أن عدد الرحلات السياحية الشهرية للأشهر يونيو ويوليو وسبتمبر هي أقل بكثير من تلك الخاصة بالشهور الأخرى من العام، لذلك من المكن من خلال هذه الملاحظة تقديم الاقتراح بالعمل على زيادة عدد الرحلات السياحية في أشهر يونيو ويوليو وسبتمبر، وليس في جميع أشهر العام.

في مُعامِلات أداة سولفر، ستكون **الخلية الهدف** (Objective Cell) هي إجمالي عدد الرحلات السياحية، وستكون خلايا المتغير هي أعداد الرحلات السياحية في الأشهر يونيو ويوليو وسبتمبر، ولن يتم إضافة قيود محددة، وسيتم تعيين قيمة خلية الهدف إلى 25,000,000.

### لاستخدام أداة سولفر بدون قيود:

- من علامة تبويب Data (بيانات)، 1 اضغط على زر Solver (سولفر). 2
- في حقل Set Objective (تحديد الهدف)، اختر الخلية C14. 3
- حدد Value of: 25000 (القيمة: 25000). 4
- من خيار By Changing Variable Cells (لتغيير الخلايا المتغيرة) اختر الخلايا C7;C8;C10. 5
- اضغط على Solve (حل). 6
- في نافذة Solver Results (نتائج سولفر)، اضغط على OK (موافق). 7
- ستُطبّق التغييرات على الخلايا المحددة. 8

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 1,653 |
| 3 | 2 | فبراير 2019 | 1,492 |
| 4 | 3 | مارس 2019 | 1,599 |
| 5 | 4 | أبريل 2019 | 1,650 |
| 6 | 5 | مايو 2019 | 2,079 |
| 7 | 6 | يونيو 2019 | 903 |
| 8 | 7 | يوليو 2019 | 750 |
| 9 | 8 | أغسطس 2019 | 2,675 |
| 10 | 9 | سبتمبر 2019 | 635 |
| 11 | 10 | أكتوبر 2019 | 1,119 |
| 12 | 11 | نوفمبر 2019 | 1,234 |
| 13 | 12 | ديسمبر 2019 | 1,735 |
| 14 | | الإجمالي | 17,526 |

C10 =SUM(C2:C13)

ينخفض عدد الرحلات السياحية الشهرية بشكل ملحوظ في أشهر يونيو ويوليو وسبتمبر، لذلك ستحاول زيادة القيم في خلايا C7 وC8 وC10 فقط من أجل تحقيق الهدف المنشود وهو رفع عدد الرحلات السياحية على مدار العام.

**الخلية الهدف** (Objective Cell) هي الخلية المُستهدفة في ورقة عملك، والتي يجب زيادة أو خفض قيمتها، أو تحديدها بقيمة معينة. هذه هي الخلية التي تحتوي على الدالة الهدف (الصيغة).

**الخلايا المتغيرة** (Variable Cells) هي الخلايا الموجودة في ورقة العمل والتي ستتغير قيمُها. هذه هي متغيرات القرار التي سيتم ضبطها بهدف الوصول إلى الحل الأمثل.

يمكنك إضافة **قيود** (Constraints) في هذه القائمة، وهي الحدود التي تفرضها على تغييرات القيم في خلايا معينة.

Solver Parameters

Set Objective: $C$14

To: Max / Min / Value Of: 25000

By Changing Variable Cells: $C$7;$C$8;$C$10

Subject to the Constraints:

Add | Change | Delete | Reset All | Load/Save

Make Unconstrained Variables Non-Negative

Select a Solving Method: GRG Nonlinear | Options

Solving Method

Select the GRG Nonlinear engine for Solver Problems that are smooth nonlinear. Select the LP Simplex engine for linear Solver Problems, and select the Evolutionary engine for Solver problems that are non-smooth.

Close | Solve | Help

Solver Results

Solver found a solution. All Constraints and optimality conditions are satisfied.

Keep Solver Solution

Restore Original Values

Reports: Answer, Sensitivity, Limits

Return to Solver Parameters Dialog

Outline Reports

OK | Cancel | Save Scenario...

**Solver found a solution. All Constraints and optimality conditions are satisfied.**

When the GRG engine is used, Solver has found at least a local optimal solution. When Simplex LP is used, this means Solver has found a global optimal solution.

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 1,653 |
| 3 | 2 | فبراير 2019 | 1,492 |
| 4 | 3 | مارس 2019 | 1,599 |
| 5 | 4 | أبريل 2019 | 1,650 |
| 6 | 5 | مايو 2019 | 2,079 |
| 7 | 6 | يونيو 2019 | 4,327 |
| 8 | 7 | يوليو 2019 | 3,109 |
| 9 | 8 | أغسطس 2019 | 2,675 |
| 10 | 9 | سبتمبر 2019 | 2,327 |
| 11 | 10 | أكتوبر 2019 | 1,119 |
| 12 | 11 | نوفمبر 2019 | 1,234 |
| 13 | 12 | ديسمبر 2019 | 1,735 |
| 14 | | الإجمالي | 25,000 |

شكل 4.30: استخدم أداة سولفر بدون قيود

## تقييم النتائج Assess the results

الآن وبعد أن قمت بتنفيذ أداة سولفر، أصبحت جاهزًا لإلقاء نظرة على نتائجها. كما تمت الإشارة إليه في بداية الدرس، فإن نتائج أداة سولفر ليست مجرد أرقام، بل هي قيم ذات معنى مخصص لموقف الدراسة المحدد، وبصفتك عالم بيانات، عليك تقييم هذه النتائج لتحديد الجدوى من القيام بالمزيد من الإجراءات. أولاً، ستُنشئ في ورقة عمل سولفر جدولاً يتضمن خمسة أعمدة وهي: الرقم التسلسلي، الشهر، الرحلات السياحية الشهرية 2019، نتائج Solver لعام 2023، والاختلاف، وبعد ذلك ستقوم بإجراء المقارنة بسهولة قبل تطبيق أداة سولفر وبعد تطبيقها.

**لحساب الاختلاف:**

- انسخ قيم الخلايا C1:C14 في ورقة العمل "2019". 1
- في ورقة عمل "Solver"، حدِّد العمود D، واضغط عليه بزر الفأرة الأيمن. 2
- ألصق القيم التي نسختها. 3
- غيِّر أسماء الأعمدة واحذف النص "2019" من جميع أشهر العمود B. 4
- أضف عمودًا باسم "الاختلاف". 5
- اكتب في الخلية E2 الصيغة "=C2-D2". 6
- نفِّذ الدالة في جميع الخلايا من E2 إلى E14، 7 ثم اضغط على زر Enter ↵.
- أصبحت ورقة "Solver" جاهزة الآن لتقييم النتائج. 8

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 1,653 |
| 3 | 2 | فبراير 2019 | 1,492 |
| 4 | 3 | مارس 2019 | 1,599 |
| 5 | 4 | أبريل 2019 | 1,650 |
| 6 | 5 | مايو 2019 | 2,079 |
| 7 | 6 | يونيو 2019 | 903 |
| 8 | 7 | يوليو 2019 | 750 |
| 9 | 8 | أغسطس 2019 | 2,675 |
| 10 | 9 | سبتمبر 2019 | 635 |
| 11 | 10 | أكتوبر 2019 | 1,119 |
| 12 | 11 | نوفمبر 2019 | 1,234 |
| 13 | 12 | ديسمبر 2019 | 1,735 |
| 14 | | الإجمالي | 17,526 |

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 1,653 |
| 3 | 2 | فبراير 2019 | 1,492 |
| 4 | 3 | مارس 2019 | 1,599 |
| 5 | 4 | أبريل 2019 | 1,650 |
| 6 | 5 | مايو 2019 | 2,079 |
| 7 | 6 | يونيو 2019 | 4,327 |
| 8 | 7 | يوليو 2019 | 3,109 |
| 9 | 8 | أغسطس 2019 | 2,675 |
| 10 | 9 | سبتمبر 2019 | 2,327 |
| 11 | 10 | أكتوبر 2019 | 1,119 |
| 12 | 11 | نوفمبر 2019 | 1,234 |
| 13 | 12 | ديسمبر 2019 | 1,735 |
| 14 | | الإجمالي | 25,000 |

قص
نسخ
خيارات اللصق:
لصق (ص)
إدراج
حذف
مسح المحتويات
تنسيق خلايا...
عرض العمود...
إخفاء
إظهار

4

5

| | الشهر | نتائج Solver لعام 2023 | الرحلات السياحية الشهرية لعام 2019 | الاختلاف |
|---|---|---|---|---|
| 1 | يناير | 1,653 | 1,653 | |
| 2 | فبراير | 1,492 | 1,492 | |
| 3 | مارس | 1,599 | 1,599 | |
| 4 | أبريل | 1,650 | 1,650 | |
| 5 | مايو | 2,079 | 2,079 | |
| 6 | يونيو | 4,327 | 903 | |
| 7 | يوليو | 3,109 | 750 | |
| 8 | أغسطس | 2,675 | 2,675 | |
| 9 | سبتمبر | 2,327 | 635 | |
| 10 | أكتوبر | 1,119 | 1,119 | |
| 11 | نوفمبر | 1,234 | 1,234 | |
| 12 | ديسمبر | 1,735 | 1,735 | |
| | الإجمالي | 25,000 | 17,526 | |

6

=C2-D2

| | الشهر | نتائج Solver لعام 2023 | الرحلات السياحية الشهرية لعام 2019 | الاختلاف |
|---|---|---|---|---|
| 1 | يناير | 1,653 | 1,653 | =C2-D2 |
| 2 | فبراير | 1,492 | 1,492 | |
| 3 | مارس | 1,599 | 1,599 | |
| 4 | أبريل | 1,650 | 1,650 | |
| 5 | مايو | 2,079 | 2,079 | |
| 6 | يونيو | 4,327 | 903 | |
| 7 | يوليو | 3,109 | 750 | |
| 8 | أغسطس | 2,675 | 2,675 | |
| 9 | سبتمبر | 2,327 | 635 | |
| 10 | أكتوبر | 1,119 | 1,119 | |
| 11 | نوفمبر | 1,234 | 1,234 | |
| 12 | ديسمبر | 1,735 | 1,735 | |
| | الإجمالي | 25,000 | 17,526 | |

7

| الاختلاف | الرحلات السياحية الشهرية لعام 2019 | نتائج Solver لعام 2023 | الشهر | |
|---|---|---|---|---|
| 0 | 1,653 | 1,653 | يناير | 1 |
| 0 | 1,492 | 1,492 | فبراير | 2 |
| 0 | 1,599 | 1,599 | مارس | 3 |
| 0 | 1,650 | 1,650 | أبريل | 4 |
| 0 | 2,079 | 2,079 | مايو | 5 |
| 3,423 | 903 | 4,327 | يونيو | 6 |
| 2,359 | 750 | 3,109 | يوليو | 7 |
| 0 | 2,675 | 2,675 | أغسطس | 8 |
| 1,692 | 635 | 2,327 | سبتمبر | 9 |
| 0 | 1,119 | 1,119 | أكتوبر | 10 |
| 0 | 1,234 | 1,234 | نوفمبر | 11 |
| 0 | 1,735 | 1,735 | ديسمبر | 12 |
| 7,474 | 17,526 | 25,000 | الإجمالي | |

8

| الاختلاف | الرحلات السياحية الشهرية لعام 2019 | نتائج Solver لعام 2023 | الشهر | |
|---|---|---|---|---|
| 0 | 1,653 | 1,653 | يناير | 1 |
| 0 | 1,492 | 1,492 | فبراير | 2 |
| 0 | 1,599 | 1,599 | مارس | 3 |
| 0 | 1,650 | 1,650 | أبريل | 4 |
| 0 | 2,079 | 2,079 | مايو | 5 |
| 3,423 | 903 | 4,327 | يونيو | 6 |
| 2,359 | 750 | 3,109 | يوليو | 7 |
| 0 | 2,675 | 2,675 | أغسطس | 8 |
| 1,692 | 635 | 2,327 | سبتمبر | 9 |
| 0 | 1,119 | 1,119 | أكتوبر | 10 |
| 0 | 1,234 | 1,234 | نوفمبر | 11 |
| 0 | 1,735 | 1,735 | ديسمبر | 12 |
| 7,474 | 17,526 | 25,000 | الإجمالي | |

شكل 4.31: حساب الاختلاف

ستقوم الآن بإلقاء نظرة على ورقة العمل وتقييم النتائج، وستلاحظ فورًا أن قيم أداة سولفر لأشهر يونيو ويوليو وسبتمبر مرتفعة للغاية.

يُقدِّم لنا إكسل سولفر اقتراحًا يمكن صياغته كالتالي: إذا أردت الوصول إلى هدف 25,000,000 رحلة سياحية في العام 2023، عليك القيام بحملات ترويجية للسياحة ليصل عدد الرحلات السياحية إلى 4,327,000 رحلة في شهر يونيو، و 3,109,000 رحلة في شهر يوليو، وأخيرًا 2,327,000 رحلة في شهر سبتمبر. واستنادًا إلى أعداد الرحلات في الشهور الأخرى، يُعَد هذا هدفًا غير واقعي، حيث أن أعلى قيمة لعدد الرحلات في العام 2019 كانت 2,675,000 رحلة سياحية، مما يعني أنه مهما كانت الحملة الترويجية التي ستنفذها ناجحة، فإنها لن تصل إلى الهدف المقترح وهو 4,327,000 رحلة في شهر يونيو، حيث أن ذلك الرقم أعلى بنسبة 160% من عدد الرحلات التي تم تسجيلها في أفضل شهر في العام 2019، كما هو موضّح في الشكل 4.32.

| الشهر | نتائج Solver لعام 2023 | الرحلات السياحية الشهرية لعام 2019 | الاختلاف |
|---|---|---|---|
| يناير | 1,653 | 1,653 | 0 |
| فبراير | 1,492 | 1,492 | 0 |
| مارس | 1,599 | 1,599 | 0 |
| أبريل | 1,650 | 1,650 | 0 |
| مايو | 2,079 | 2,079 | 0 |
| يونيو | 4,327 | 903 | 3,423 |
| يوليو | 3,109 | 750 | 2,359 |
| أغسطس | 2,675 | 2,675 | 0 |
| سبتمبر | 2,327 | 635 | 1,692 |
| أكتوبر | 1,119 | 1,119 | 0 |
| نوفمبر | 1,234 | 1,234 | 0 |
| ديسمبر | 1,735 | 1,735 | 0 |
| الإجمالي | 25,000 | 17,526 | 7,474 |

عدد الرحلات السياحية لأشهر يونيو ويوليو وسبتمبر التي تم إنشاؤها بواسطة أداة سولفر مرتفع جدًا مقارنةً بأعداد الرحلات في جميع الشهور الأخرى.

عدد الرحلات السياحية لأشهر يونيو ويوليو وسبتمبر 2019.

شكل 4.32: عدد الرحلات السياحية لعام 2019 والبيانات بعد تطبيق أداة سولفر

ولأن النتائج الأولية بعد تطبيق أداة سولفر في المثال السابق ليست واقعية، يتعين عليك تنفيذ تشغيل أداة سولفر من جديد، ولكن مع تحديد بعض القيود لأجل الحصول على أهداف أكثر واقعية. على سبيل المثال، يمكنك ضبط أداة سولفر لحساب القيم دون قيد لجميع الشهور باستثناء أشهر يونيو ويوليو وسبتمبر، والتي يجب تعيين قيد لقيمها لتجنُّب الحصول على نتائج غير واقعية، ويمكن تحقيق ذلك باستخدام متوسط قيم البيانات الموجودة. ستسمح لأداة سولفر بحساب القيم لجميع الشهور، ولكن بإضافة قيود تُحدد أن الأعداد المستهدفة لأشهر يونيو ويوليو وسبتمبر يجب أن تكون أعلى من المتوسط الشهري للعام 2019.

للعمل على ملف إكسل نفسه، ستُنشئ نسخة أخرى من ورقة العمل "2019" باسم "Solver (constraints)" وستقوم بتشغيل أداة سولفر مرة أخرى، ولكن هذه المرة بقيود.



## Calculate the average حساب المتوسط

يتم حساب **المتوسط** (Average) أو ما يسمى **بالوسط الحسابي** (Arithmetic mean) لمجموعة من الأرقام بجمعها ثم قسمة الناتج على عدد هذه الأرقام. في الحالة التالية، ستجمع إجمالي عدد الرحلات السياحية الشهرية ثم تقسم الناتج على 12.

**لحساب المتوسط:**

- في ملف إكسل جديد، أنشئ ورقة عمل جديدة وأطلق عليها اسم "Solver (constraints)". 1
- في الخلية B15، اكتب "المتوسط". 2
- في الخلية C15، اكتب الصيغة "=average(C2:C13)". 3
- اضغط على الزر **Enter ↵** وسيظهر المتوسط في الخلية C15. 4

**الوسط الحسابي:**

يتم حساب المُعدَّل في الإحصاء الوصفي عن طريق جمع النتائج معًا ثم قسمة الإجمالي على عدد هذه النتائج.

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 1,653 |
| 3 | 2 | فبراير 2019 | 1,492 |
| 4 | 3 | مارس 2019 | 1,599 |
| 5 | 4 | أبريل 2019 | 1,650 |
| 6 | 5 | مايو 2019 | 2,079 |
| 7 | 6 | يونيو 2019 | 903 |
| 8 | 7 | يوليو 2019 | 750 |
| 9 | 8 | أغسطس 2019 | 2,675 |
| 10 | 9 | سبتمبر 2019 | 635 |
| 11 | 10 | أكتوبر 2019 | 1,119 |
| 12 | 11 | نوفمبر 2019 | 1,234 |
| 13 | 12 | ديسمبر 2019 | 1,735 |
| 14 | | الإجمالي | 17,526 |
| 15 | | المتوسط | =average(C2:C13) |

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 1,653 |
| 3 | 2 | فبراير 2019 | 1,492 |
| 4 | 3 | مارس 2019 | 1,599 |
| 5 | 4 | أبريل 2019 | 1,650 |
| 6 | 5 | مايو 2019 | 2,079 |
| 7 | 6 | يونيو 2019 | 903 |
| 8 | 7 | يوليو 2019 | 750 |
| 9 | 8 | أغسطس 2019 | 2,675 |
| 10 | 9 | سبتمبر 2019 | 635 |
| 11 | 10 | أكتوبر 2019 | 1,119 |
| 12 | 11 | نوفمبر 2019 | 1,234 |
| 13 | 12 | ديسمبر 2019 | 1,735 |
| 14 | | الإجمالي | 17,526 |
| 15 | | المتوسط | 1,460 |

Solver (constraints) | Solver | 2019

شكل 4.33: حساب المتوسط

## استخدام أداة سولفر مع قيود Solver with constraints

الآن وبعد أن حسبت المتوسط الشهري لعدد الرحلات السياحية لعام 2019، ستستخدم أداة سولفر من جديد بقيود، بحيث تكون الخلية الهدف في مُعاملات أداة سولفر هي إجمالي عدد الرحلات السياحية، وستكون الخلايا المتغيرة هي أعداد الرحلات السياحية لجميع الشهور. ستتم إضافة قيود محددة، مع تعيين القيم المقدرة لأداة سولفر لأشهر يونيو ويوليو وسبتمبر لتكون أكبر من أو تساوي متوسط عدد الرحلات في العام 2019 وهو 1,460,000 ألف رحلة سياحية. وأخيرًا، سيتم تعيين قيمة الخلية الهدف مرةً أخرى إلى 25,000,000.

**القيد (Constraint):**

القيد هو تحديد معين لما يُمكن تضمينه في مجموعة البيانات أو البيانات الوصفية، أو تحديد لمجموعة من مفاتيح البيانات التي يمكن إرفاق سمات معينة بها (وتُحدد بواسطة هيكل البيانات).

**لاستخدام أداة سولفر بقيود:**

< من علامة تبويب Data (بيانات)، 1 اضغط على Solver (سولفر). 2

< من حقل Set Objective (تعيين الهدف) اختر الخلية C14. 3

< اختر Value of: 25000 (القيمة:25000). 4

< من By Changing Variable Cells (لتغيير خلايا المتغير)، حدِّد الخلايا C2:C13. 5

< اضغط على Add (إضافة) لإضافة قيد. 6

| الشهر | الرحلات السياحية |
| --- | --- |
| يناير 2019 | 1,653 |
| فبراير 2019 | 1,492 |
| مارس 2019 | 1,599 |
| أبريل 2019 | 1,650 |
| مايو 2019 | 2,079 |
| يونيو 2019 | 903 |
| يوليو 2019 | 750 |
| أغسطس 2019 | 2,675 |
| سبتمبر 2019 | 635 |
| أكتوبر 2019 | 1,119 |
| نوفمبر 2019 | 1,234 |
| ديسمبر 2019 | 1,735 |
| الإجمالي | 17,526 |
| المتوسط | 1,460 |

شكل 4.34: استخدام أداة سولفر بقيود

## لإعداد القيود:

> من مربع Cell reference (مرجع الخلية)، حدِّد الخلية C7. 1
> اختر الرمز <=. 2
> ضَع الرقم 1460 في صندوق Constraint (القيد). 3
> اضغط على Add (إضافة). 4
> اضبُط بقية قيود الخلايا C8 و C10 واضغط على OK (موافق). 5

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 1,653 |
| 3 | 2 | فبراير 2019 | 1,492 |
| 4 | 3 | مارس 2019 | 1,599 |
| 5 | 4 | أبريل 2019 | 1,650 |
| 6 | 5 | مايو 2019 | 2,079 |
| 7 | 6 | يونيو 2019 | 903 |
| 8 | 7 | يوليو 2019 | 750 |
| 9 | 8 | أغسطس 2019 | 2,675 |
| 10 | 9 | سبتمبر 2019 | 635 |
| 11 | 10 | أكتوبر 2019 | 1,119 |
| 12 | 11 | نوفمبر 2019 | 1,234 |
| 13 | 12 | ديسمبر 2019 | 1,735 |
| 14 | | الإجمالي | 17,526 |
| 15 | | المتوسط | 1,460 |

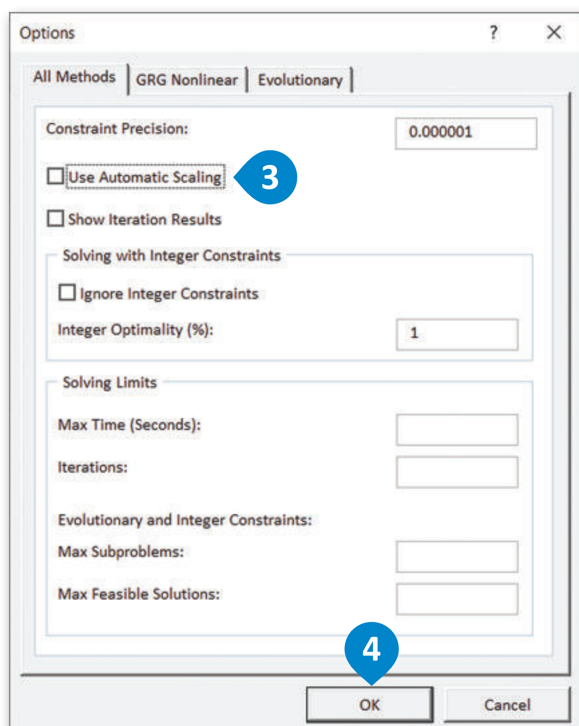
شكل 4.35: إعداد القيود.

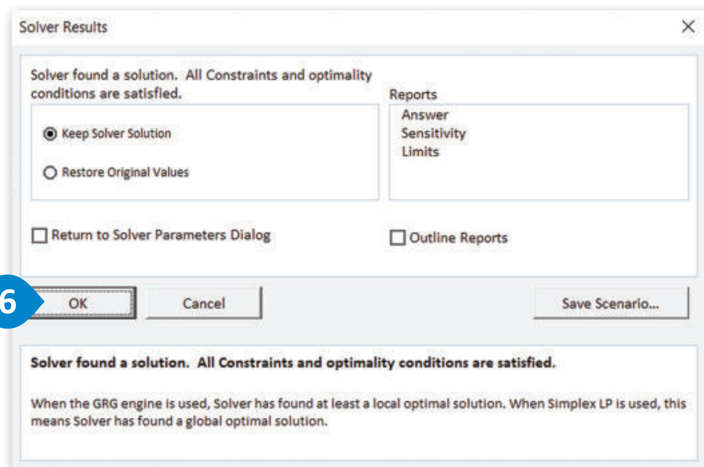
Add Constraint
Cell Reference: $C$7 | <= | Constraint: 1460
Cancel | Add | OK

## لتعيين مُعاملات أداة سولفر:

> تأكَّد من عرض قائمة القيود بشكل صحيح في مربع "Subject to the Constraints" (تخضع للقيود). 1
> اضغط على Options (خيارات). 2
> في نافذة الخيارات، ألغِ تحديد خيار Use Automatic Scaling (استخدام القياس التلقائي). 3
> اضغط على OK (موافق). 4
> في نافذة Solver Parameters (مُعاملات Solver)، اضغط على Solve (حل). 5
> في نافذة Solver Results (نتائج Solver)، اضغط على OK (موافق). 6
> ستُطبق التغييرات في الخلايا المحددة. 7

Solver Parameters
Set Objective: $C$14
To: Max / Min / Value Of: 25000
By Changing Variable Cells: $C$2:$C$13
Subject to the Constraints:
$C$10 <= 1460
$C$7 <= 1460
$C$8 <= 1460
Add | Change | Delete | Reset All | Load/Save
Make Unconstrained Variables Non-Negative
Select a Solving Method: GRG Nonlinear | Options
Solving Method
Select the GRG Nonlinear engine for Solver Problems that are smooth nonlinear. Select the LP Simplex engine for linear Solver Problems, and select the Evolutionary engine for Solver problems that are non-smooth.
Help | Solve | Close




C14 =SUM(C2:C13)

| | A | B | C |
|---|---|---|---|
| 1 | | الشهر | الرحلات السياحية |
| 2 | 1 | يناير 2019 | 2,156 |
| 3 | 2 | فبراير 2019 | 1,901 |
| 4 | 3 | مارس 2019 | 2,069 |
| 5 | 4 | أبريل 2019 | 2,151 |
| 6 | 5 | مايو 2019 | 2,874 |
| 7 | 6 | يونيو 2019 | 1,610 |
| 8 | 7 | يوليو 2019 | 1,563 |
| 9 | 8 | أغسطس 2019 | 3,989 |
| 10 | 9 | سبتمبر 2019 | 1,534 |
| 11 | 10 | أكتوبر 2019 | 1,349 |
| 12 | 11 | نوفمبر 2019 | 1,514 |
| 13 | 12 | ديسمبر 2019 | 2,289 |
| 14 | | الإجمالي | 25,000 |
| 15 | | المتوسط | 2,083 |

ستظهر التغييرات في الخلايا C2: C13.

7

القيم الموجودة في الخلايا C7 و C8 و C10 أعلى من المتوسط (1,460).

شكل 4.36: تعيين مُعامِلات أداة سولفر

## تقييم سولفر باستخدام نتائج القيود Assess the Solver with constraints results

الآن وبعد أن استخدمت أداة سولفر بقيود، يُمكنك إنشاء جدول مُكوَّن من الأعمدة الخمسة التالية مرة أخرى: الرقم التسلسلي، والشهر، والرحلات السياحية الشهرية لعام 2019، ونتائج (Constraints) Solver لعام 2023، والاختلاف، وبهدف مقارنة البيانات بسهولة قبل استخدام أداة سولفر وبعدها. من خلال إلقاء نظرة على النتائج، يمكنك هذه المرة ملاحظة أن إكسل بمساعدة سولفر يُقدِّم اقتراحًا بأنه إذا أردت الوصول إلى هدف 25,000,000 زيارة سياحية في العام 2023، فيجب تنفيذ حملتك الترويجية للسياحة بحيث تتم زيادة أعداد الرحلات السياحية لجميع شهور السنة، كما يجب أن تكون الحملة الترويجية للسياحة أكثر شمولية على مدار العام ولا تقتصر فقط على الأشهر يونيو ويوليو وسبتمبر حيث وُجِدت الأرقام ذات الإشكالية في البداية. تشير نتائج سولفر أيضًا إلى أن حملتك السياحية يجب أن تُركِّز على زيادة عدد الزيارات، ولكن بصورة واقعية.

| الشهر | نتائج (Constraints) Solver لعام 2023 | الرحلات السياحية الشهرية لعام 2019 | الاختلاف |
|---|---|---|---|
| يناير | 2,156 | 1,653 | 503 |
| فبراير | 1,901 | 1,492 | 409 |
| مارس | 2,069 | 1,599 | 470 |
| أبريل | 2,151 | 1,650 | 501 |
| مايو | 2,874 | 2,079 | 795 |
| يونيو | 1,610 | 903 | 707 |
| يوليو | 1,563 | 750 | 813 |
| أغسطس | 3,989 | 2,675 | 1,314 |
| سبتمبر | 1,534 | 635 | 899 |
| أكتوبر | 1,349 | 1,119 | 230 |
| نوفمبر | 1,514 | 1,234 | 280 |
| ديسمبر | 2,289 | 1,735 | 554 |
| الإجمالي | 25,000 | 17,526 | 7,474 |
| المتوسط | 2,083 | | |

شكل 4.37: أعداد الرحلات السياحية لعام 2019 والبيانات بعد تنفيذ سولفر بقيود

أعداد الرحلات السياحية لأشهر يونيو ويوليو وسبتمبر 2019.

أعداد الرحلات السياحية للأشهر يونيو ويوليو وسبتمبر التي تم إنشاؤها بواسطة أداة سولفر مع القيود. هذه القيم واقعية وستكون مفيدة في اتخاذ القرارات المستقبلية.

في الختام، تُظهر نتائج سولفر مع قيود بأنه يجب تصميم حملة ترويجية شاملة للسياحة لتعزيز عدد الرحلات السياحية في كل شهر من شهور العام، مع هدف يتراوح بين 500,000 إلى 1,000,000 رحلة سياحية شهريًا. يمكن لوكالات السياحة والسفر الاستفادة من هذه الاقتراحات ووضع استراتيجيات لتعزيز الحملات الترويجية للسياحة في المملكة العربية السعودية على مدار العام، مع التركيز بشكل خاص على الأشهر يونيو ويوليو وسبتمبر. يُمكن على سبيل المثال تقديم خصومات على تذاكر الطيران أو تنظيم واستقطاب رحلات بحرية، وتنظيم المهرجانات المختلفة لجذب المزيد من السياح خلال هذه الأشهر الثلاثة.



# تمرينات

1

| حدِّد الجملة الصحيحة والجملة الخاطئة فيما يلي: | صحيحة | خاطئة |
|---|---|---|
| 1. تعتبر أداة إكسل أداة أداة في إكسل تساعدك على تحسين النماذج. | ○ | ○ |
| 2. يعتبر تصميم الحملة الترويجية للسياحة مشكلة من مشاكل التحسين. | ○ | ○ |
| 3. ليس من الضروري صياغة المشكلة قيد الدراسة مسبقًا. | ○ | ○ |
| 4. يتم ضبط الخلية الهدف إلى قيمة محددة دائمًا. | ○ | ○ |
| 5. نادرًا ما يتم استخدام أداة سولفر بقيود. | ○ | ○ |
| 6. يُعتبر تقييم نتائج سولفر جزءًا من عملية التحسين المستمر. | ○ | ○ |
| 7. من المهم مقارنة قيم البيانات السابقة بالقيم المتوقعة من أجل الوصول إلى استنتاجات أفضل. | ○ | ○ |
| 8. يجب ألا تتجاوز نتائج سولفر متوسط القيم المحددة. | ○ | ○ |
| 9. يتم اختيار الخلايا المتغيرة بناءً على الظاهرة أو المشكلة قيد الدراسة. | ○ | ○ |
| 10. يجب عدم وجود ارتباط بين خلية الهدف وخلايا المتغير. | ○ | ○ |



**2** قارن بين استخدام أداة سولفر بدون قيود واستخدامها بقيود، ثم اذكر اثنين من الاختلافات الأساسية بينهما.

________________________________________________

________________________________________________

________________________________________________

________________________________________________

________________________________________________

________________________________________________

________________________________________________

________________________________________________

**3** قُمْ بزيارة صفحة بيانات السياحة التابعة لوزارة السياحة السعودية في البوابة الوطنية للبيانات المفتوحة (https://data.gov.sa)، ثم قُمْ بتنزيل بيانات الرحلات السياحية لعامي 2017 و 2018. استخدم أداة سولفر لتصميم حملتك السياحية لعام 2023.

هل يُمكن استخدام هذه البيانات لهذا الغرض؟ وضِّح إجابتك.

________________________________________________

________________________________________________

________________________________________________

________________________________________________

________________________________________________

________________________________________________

________________________________________________

________________________________________________

________________________________________________

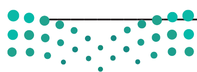
________________________________________________

________________________________________________

________________________________________________

________________________________________________

________________________________________________

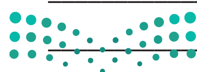


4 قَيّم نتائج أداة سولفر التي استخدمتها في التدريب الثالث، واشرح ما إذا كانت واقعية أم لا؟ واذكر الحلول الأخرى التي لديك من أجل الوصول إلى نتائج أفضل وسبب اختيارك لها.

________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________

5 قيِّم نتائج التدريب الرابع وقارنها بالنتائج الموضحة في الدرس، ثم اذكر النتائج التي تجدها مناسبة للقيام بحملتك الترويجية للسياحة.

________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________
________________________________________



# المشروع

1

افترض أنك تعمل كَوكيل للسفريات وترغب في تنفيذ حملة ترويجية للسياحة للمملكة العربية السعودية لعام 2024، حيث تهدف إلى رفع عددٍ الرحلات السياحية إلى 50,000,000.

2

**ويتعين عليك:**

- تحميل بيانات الرحلات السياحية لعام 2018 من صفحة بيانات السياحة التابعة لوزارة السياحة السعودية في البوابة الوطنية للبيانات المفتوحة (https://data.gov.sa).
- إنشاء توقُّع لهذه البيانات لعام 2024.
- تقييم النتائج المتوقّعة من أجل صياغة مشكلة التحسين.
- استخدم أداة إكسل سولفر (Excel Solver) للحصول على معلومات حول كيفية تصميم حملتك الترويجية للسياحة.
- تقديم اقتراحات لحملة ترويجية للسياحة بناءً على نتائج أداة إكسل سولفر.

3

قم بإعداد عرضًا تقديميًا على برنامج مايكروسوفت باوربوينت باستخدام المُخططات أو الرسوم البيانية المناسبة في إكسل والتي تُوضِّح توقُّعاتك ونتائج أداة إكسل سولفر، ثم اشرح مخططاتك وقدِّم اقتراحاتك للحملة الترويجية للسياحة.





# ماذا تعلمت

- مصطلح النمذجة التنبؤية.
- الاختلاف بين النمذجة المُعامِلية وغير المُعامِلية.
- الأنواع المختلفة للنماذج التنبؤية.
- عملية إنشاء نموذج تنبؤي.
- ميزات وتحديات النمذجة التنبؤية.
- تطبيقات النمذجة التنبؤية.
- مصطلح التوقُّع.
- الأنواع المختلفة لمخططات التوقُّع.
- خطوات تطبيق توقُّع على بيانات محددة.
- مفهوم نطاق الثقة.
- مصطلح التحسين وإيجاد الحلول المثالية للمشكلة.
- خطوات إجراء التحسينات باستخدام أداة إكسل سولفر.
- خطوات تقييم نتائج أداة إكسل سولفر.



## المصطلحات الرئيسة

| | | | |
|---|---|---|---|
| Classification Model | نموذج التصنيف | Lower Confidence Bound | الحد الأدنى للثقة |
| Clustered Column Chart | مخطط عمودي متفاوت المسافات | Model Formulation | تكوين النموذج |
| Clustering Model | نموذج التجميع | Neural Network | شبكة عصبية |
| Column Chart | مخطط عمودي | Non-Parametric | غير مُعامِلي |
| Confidence Interval | نطاق الثقة | Objective Cell | الخلية الهدف |
| Constraints | القيود | Optimization Problem | مشكلة التحسين |
| Data Collection | جمع البيانات | Outlier Detection Model | نموذج اكتشاف القيم الشاذة |
| Data Purification | تنقية البيانات | Parametric | مُعامِلي |
| Data Transformation | تحويل البيانات | Predictive Data Modeling | نمذجة البيانات التنبؤية |
| Decision Tree | شجرة القرار | Prophet Model | نموذج بروفيت |
| Excel Solver | أداة إكسل سولفر | Quality Enhancement | تحسين الجودة |
| Forecast | توقُّع | Risk Assessment | تقييم المخاطر |
| Forecast Model | نموذج التوقُّع | Stacked Column Chart | مخطط عمودي مكدس |
| General Linear Model | نموذج خطي عام | Time Series Model | نموذج السلاسل الزمنية |
| Gradient Boosted Model | نموذج التعزيز الاشتقاقي | Upper Confidence Bound | الحد الأعلى للثقة |
| Line Chart | مخطط خطي | Variable Cells | خلايا المتغير |
| Linear Regression | الانحدار الخطي | | |

# متطلب البرمجة بلغة بايثون

تعد البرمجة أحد أهم المهارات التي ينبغي اكتسابها من قبل الطلاب الملتحقين بمسار علوم الحاسب والهندسة حيث تعد متطلب لعدد من المناهج في هذا المسار ومنها منهجي الهندسة وعلم البيانات. لتسهيل اكتساب الطالب لأساسيات البرمجة بلغة البايثون، فقد تم تصميم المحتوى التالي الذي يمكن الوصول إليه بمسح رمز الاستجابة السريع الخاص بكل موضوع. وينصح الطالب بوضع خطة زمنية لإتمام الاطلاع على هذه الوحدات ويمكن الاسترشاد بالمدة الزمنية المقترحة لكل وحدة كما يمكن للطالب وضع علامة (✓) لتعليم الوحدات التي أتمها.

| الوحدة | المدة الزمنية المقترحة | رمز الاستجابة السريع | هل أتممت الوحدة؟ |
|---|---|---|---|
| 1 . مقدمة في البايثون<br>Introduction to Python | يوم واحد | | |
| 2. المدخلات والمخرجات والعمليات الحسابية<br>Input-Output and Mathematical Operations | يوم واحد | | |
| 3 . الجمل الشرطية<br>Conditional Statements | يومان | | |



| الوحدة | المدة الزمنية المقترحة | رمز الاستجابة السريع | هل أتممت الوحدة؟ |
|---|---|---|---|
| 4. التكرارات والدوال<br>Loops and Functions | يومان | | |
| 5 . القوائم وصفوف البيانات والمكتبات البرمجية<br>Lists, Tuples and Python Libraries | أسبوع | | |
| 6. القواميس والقوائم المتداخلة وملفات البيانات<br>Dictionary, Nested Lists and Data Files | أسبوع | | |
| 7 . هياكل البيانات المتقدمة ودوال الاستدعاء الذاتي<br>Advanced Data Structures and Recursion | أسبوعان | | |
| 8 . مقدمة في البرمجة الكائنية<br>Introduction to Object Oriented Programming | أسبوعان | | |

